THE BELL SYSTEM

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

VOLUME XXXIX JULY 1960 NUMBER 4

The Theory and Design of Chirp Radars J. R. KLAUDER, A. C. PRICE, S. DARLINGTON AND W. J. ALBERSHEIM 745

The Design of Radar Signals Having Both High Range Resolution and High Velocity Resolution J. R. KLAUDER 809

The TJ Radio Relay System J. GAMMIE AND S. D. HATHAWAY 881

Evaluation of Message Circuit Noise A. J. AIKENS AND D. A. LEWINSKI 879

A New Measuring Set for Message Circuit Noise W. T. COCHRAN AND D. A. LEWINSKI 911

Impurity Redistribution and Junction Formation in Silicon by Thermal Oxidation M. M. ATALLA AND E. TANNENBAUM 933

Synthesis of Driving-Point Impedances with Active *RC* Networks I. W. SANDBERG 947

Pressure at the Interface of Inner Conductor and Dielectric of Armorless Ocean Cable P. G. BHUTA 963

The Construction of Missile Guidance Codes Resistant to Random Interference A. R. BOKLER 973

Human Factors Engineering Studies of the Design and Use of Pushbutton Telephone Sets R. L. DEININGER 995

Passage Effects in Paramagnetic Resonance Experiments M. WEGER 1013

Recent Bell System Monographs 1113

Contributors to This Issue 1121

COPYRIGHT 1960 AMERICAN TELEPHONE AND TELEGRAPH COMPANY

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

H. I. ROMNES, *President, Western Electric Company*

J. B. FISK, *President, Bell Telephone Laboratories*

E. J. McNEELY, *Executive Vice President, American Telephone and Telegraph Company*

EDITORIAL COMMITTEE

A. C. DICKIESON, *Chairman*	K. E. GOULD
S. E. BRILLHART	E. I. GREEN
A. J. BUSCH	G. GRISWOLD, JR.
L. R. COOK	J. R. PIERCE
R. L. DIETZOLD	M. SPARKS
J. H. FELKER	W. O. TURNER

EDITORIAL STAFF

W. D. BULLOCH, *Editor*
R. M. FOSTER, JR., *Assistant Editor*
C. POLOGE, *Production Editor*
J. T. MYSAK, *Technical Illustrations*
T. N. POPE, *Circulation Manager*

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y. F. R. Kappel, President; S. Whitney Landon, Secretary; L. Chester May, Treasurer. Subscriptions are accepted at $5.00 per year. Single copies $1.25 each. Foreign postage is $1.08 per year or 18 cents per copy. Printed in U.S.A.

THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XXXIX JULY 1960 NUMBER 4

Copyright 1960, American Telephone and Telegraph Company

The Theory and Design of Chirp Radars

By J. R. KLAUDER, A. C. PRICE,
S. DARLINGTON and W. J. ALBERSHEIM

(Manuscript received April 5, 1960)

A new radar technique has been developed that provides a solution for the conflicting requirements of simultaneous long-range and high-resolution performance in radar systems. This technique, called Chirp at Bell Telephone Laboratories, recognizes that resolution depends on the transmitted pulse bandwidth. A long high-duty-factor transmitted pulse, with suitable modulation (linear frequency modulation in the case of Chirp), which covers a frequency interval many times the inherent bandwidth of the envelope, is employed. The receiver is designed to make optimum use of the additional signal bandwidth. This paper contains many of the important analytical methods required for the design of a Chirp radar system. The details of two signal generation methods are considered and the resulting signal waveforms and power spectra are calculated. The required receiver characteristics are derived and the receiver output waveforms are presented. The time-bandwidth product is introduced and related to the effective increase in the performance of Chirp systems. The concept of a matched filter is presented and used as a reference standard in receiver design. The effect of amplitude and phase distortion is analyzed by the method of paired echoes. One consequence of the signal design is the presence of time side lobes on the receiver output pulse analogous to the spatial side lobes in antenna theory. A method to reduce the time side lobes by weighting the pulse energy spectrum is explained in terms of paired echoes. The weighting process is described, and calculated pulse envelopes, weighting network characteristics and dele-

terious effects are presented. The effects of quadratic phase distortion are analyzed and the resultant pulse envelopes are presented. The receiver response characteristics in the presence of Doppler-shifted signals from moving targets are examined. Schematic ambiguity diagrams are presented for current signal designs.

Foreword

A recent declassification makes it possible to publish information concerning an important advance in military electronics that has been pursued in classified work at Bell Telephone Laboratories since 1951.

Traditionally, warfare has been a continuing competition between armament and firepower. This competition has carried over to the electronics of warfare. In particular, radar has been hard put to keep pace with the requirements determined by faster and higher-flying aircraft and missiles and the various radar countermeasures that they require. Therefore, anything that enhances the capability of radar, either in range, range resolution or rate of acquisition has always been most welcome.

Simple pulsed radar is limited in range by the average power radiated, in resolution by the pulse length and in acquisition time by the beam width. The design of any radar involves a compromise among these three factors—range, resolution and speed. Anything that eases this compromise is of great interest.

During most of the war, the principal emphasis in development of radar techniques and components was centered on increasing range and range resolution performance by increasing transmitter peak power and reducing pulse length. Thus, shortly after the end of the war, transmitting tubes were in existence with average power capabilities in excess of 100 times that which could be used in straightforward short-pulse operation. At about this time a number of investigators realized that targets separated in range by a distance ΔR could be resolved by using an appropriately modulated pulse provided that $c/2\Delta f \geqq \Delta R$, where c is the velocity of propagation and Δf is the width of the frequency spectrum of the transmitted pulse. The familiar case of short-pulse transmission where the free-space wave train is of the order of twice the distance between targets to be resolved satisfies this condition. The important point is that this criterion can also be satisfied with a variety of other signal waveforms. Thus, a linearly frequency-modulated pulse with frequency span set by the resolution required and with duration set by the energy required in the pulse for range performance is at once suggested. By transmitting a modulated pulse, it is possible to obtain range resolutions

that are smaller than the transmitted pulse length by very large factors. The range resolution is inherently limited to approximately the reciprocal of the signal bandwidth, and one can (in retrospect) imagine many methods of pulse modulation that utilize a given bandwidth as efficiently as does a short pulse, but that do not involve the high peak powers of a very short pulse.

Such considerations led to early theoretical and experimental work on various types of radar signals. Some of this work was concerned primarily with frequency-modulated pulses. At Bell Telephone Laboratories this followed ideas that were first proposed in 1947 by Darlington in connection with waveguide transmission, and were later covered by a patent.[1] This approach is also to be found in the work of Hüttman,[2] Sproule and Hughes,[3] Cauer,[4] and Dicke.[5] * The early experimental work at Bell Telephone Laboratories by A. F. Dietrich and O. E. De Lange and by W. J. Albersheim in 1951 utilized a reflex klystron to obtain the frequency-modulated or "Chirped" † transmitted signal. The receiver was modified to obtain the resolution compatible with the transmitted bandwidth.

Other work by C. C. Cutler envisaged more elaborate signals which included frequency-modulated pulses as a special case. Analytical work by A. W. Schelling clarified the ideas involved and clearly pointed up the problems of "side lobes in range" which arise from pulse and spectrum shaping.

The fact that the pulse compression idea, so valuable to radar, grew out of the work of a mathematician on a problem in waveguide communication is an illustration of how each branch of our complex technology supports the other.

Fundamental work in this field has been supported since the Spring of 1955 by the Weapons Guidance Laboratory of Wright Air Development Division in Dayton, Ohio.

Today, the pulse compression technique is in use in a variety of very important radar applications. Through its use, these radars transmit a pulse with about 100 times the energy of a short pulse with equivalent resolution and peak power. The paper that follows covers in detail the design and analysis necessary in applying this technique, and a second paper by Klauder[7] covers an aspect of this technique that is still evolving: the synthesis of a signal function to provide a specified ambiguity diagram. It is interesting to note that here, too, the technique has been

* An investigation based on this patent is reported by Cook.[6]

† The appellation "Chirp" was first used by B. M. Oliver in an internal Bell Laboratories Memorandum, "Not with a Bang, but a Chirp", in 1951.

advanced because Klauder recognized the similarity between the processes of synthesis of a signal function to get a specified ambiguity diagram and a method for the generation of quantum-mechanical wave functions developed by E. Wigner.

I. INTRODUCTION

The specifications of modern, highly complex communication and defense systems continually require more powerful and more sophisticated radar or radar-like systems. Such radars must be capable of detecting very small targets at greater ranges and with greater range and angular resolution than ever before. To satisfy these demands, unusual innovations in radar techniques have been required. This paper is devoted to some of the theoretical and design considerations of one such innovation, the so-called "Chirp" system, which has proved highly successful.

A familiar and intuitive figure of merit on the performance of a radar is provided by the average transmitted power: the higher the transmitted power, the greater the range of detectable targets, other things being equal. Suppose one regards the pulse repetition frequency (PRF) as being essentially fixed and seeks ways to profitably increase the transmitted power in each pulse. Clearly, an amplification of each transmitted signal constitutes the natural way to increase this power. However, the usefulness of this approach is fixed by fundamental *equipment limitations*, which frequently take the form of component breakdown. Furthermore, this presents such a serious limitation that it is necessary to seek additional means of raising the transmitted power to satisfy the requirements of modern systems.

To accomplish this it is necessary to use the only "dimension" remaining available: the time. One can surely increase the average power if the duration of each pulse is increased. Again, this technique has its limitations. First, there will be a maximum obtainable pulse width limited typically by *duty factor* considerations, i.e., the fraction of time the transmitting tubes will actually be in operation. However, there is a second limitation dictated by the desired *range resolution* of the over-all system, which determines the ability to separate multiple targets clustered closely together. In numerous present and projected systems it is an important fact that the pulse-width limitation imposed by the desired range resolution may be *several orders of magnitude less* than the limitation imposed by the present "state-of-the-art" in high duty-factor tubes. Thus, high-power, high-resolution radars are, in principle, *not* limited in a fundamental way by equipment breakdown. Rather, they are affected by poor "signal design". What is required is a transmitted

signal that combines the large amplitude and long pulse width available with existing apparatus, but retains the range resolution capabilities inherent in a pulse of much shorter duration.

One solution to this problem was recognized by Darlington.[1] Although his precise technique will not be pursued in detail, it does provide an intuitive picture of the operation of an actual Chirp system. The following section presents Darlington's model and a survey of the remaining sections of the present paper.

II. DISCUSSION OF A SIMPLE CHIRP MODEL; SURVEY AND SUMMARY OF REMAINING SECTIONS

It is the large-frequency content, or bandwidth, of a short radar pulse that accounts for its high resolution capabilities. It follows, as a consequence of Fourier analysis, that a long pulse of constant carrier frequency contains a narrow bandwidth and, therefore, possesses poor resolution properties. However, the spectrum of this long signal can be significantly broadened by introducing modulation. To utilize the transmitting tubes efficiently, this modulation must take the form of *frequency modulation* (FM). By this method one can introduce the frequency-spread characteristic of a short pulse within the envelope of a long-duration signal. The linearity of a radar system permits one to realize the potential of this shorter pulse by a suitable phase equalization in the radar receiver. Darlington's model for such a system is illustrated in Fig. 1. The transmitted signal consists of a sequence of adjacent pulses each possessing a unique carrier frequency, f_n [see Fig. 1(a)]. To realize the short pulse potential, one imagines that the received signal is passed

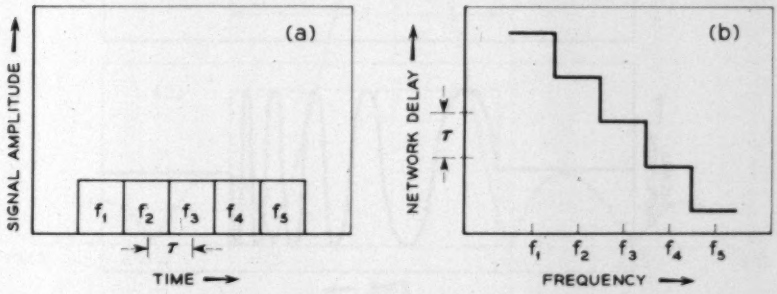


Fig. 1 — (a) Model Chirp radar signal composed of five adjacent pulses each possessing a unique frequency. (b) Suitable delay equalizer for the signal in (a). The output of this network is qualitatively a pulse of increased amplitude with a pulse width of τ seconds. The network is illustrated for the particular case where $f_n < f_{n+1}$.

through a network possessing a delay versus frequency character, as illustrated in Fig. 1(b). Although the frequency f_1 is the first to be received, it is just this frequency that is delayed the longest amount by the network. In this manner each pulse of distinct carrier frequency is made to "wait" until the highest-frequency component arrives, whereupon all the short pulses emerge simultaneously. Thus, following the delay equalization, the original signal of Fig. 1(a) will be compressed in time, and, by energy conservation, the collapsed signal will be necessarily increased in amplitude.

For several practical as well as theoretical reasons the signal chosen for the Chirp system is characterized by a *linear* FM illustrated, for example, by Fig. 2. This important case consists of a rectangular envelope of T seconds duration [see Fig. 2(a)]. Within this envelope the

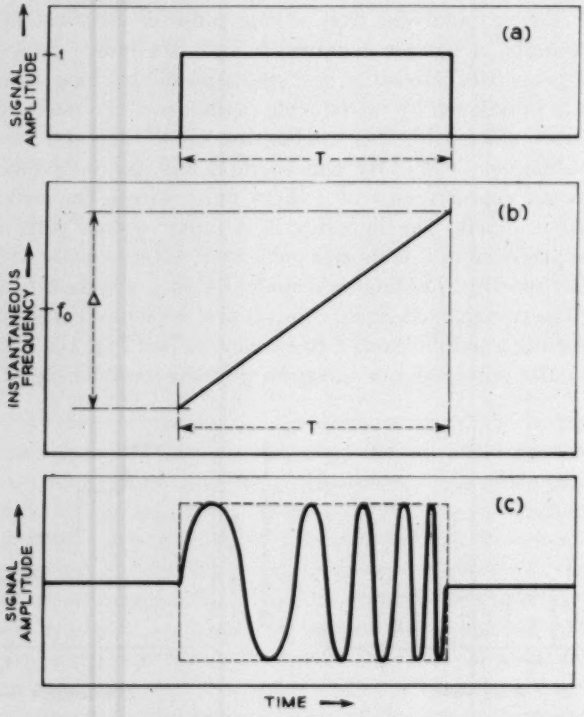


Fig. 2 — (a) Ideal envelope of actual Chirp signal, of T seconds duration and chosen to be of unit amplitude. (b) Instantaneous frequency vs. time characteristic of Chirp signal; a band of frequencies, Δ, centered at f_0 is linearly swept during the pulse duration. (c) Schematic diagram of a signal having the properties indicated in (a) and (b).

instantaneous frequency is modulated in a linear manner covering a band of frequencies Δ, centered at a frequency f_0 [see Fig. 2(b)]. Fig. 2(c) schematically illustrates a signal possessing a linear FM. It is clear that this signal is a limiting form of the one illustrated in Fig. 1(a). A suitable delay-equalizing network for the signal with linear FM is given, therefore, by a limiting form of the equalizer in Fig. 1(b). The delay characteristic of this limiting network is shown in Fig. 3(a). The response of the delay network to the rectangular pulse with linear FM is discussed in Section 3.1 of this paper. This response envelope is given analytically by the absolute value of $\sqrt{D}[(\sin \pi\Delta t)/(\pi\Delta t)]$ and is

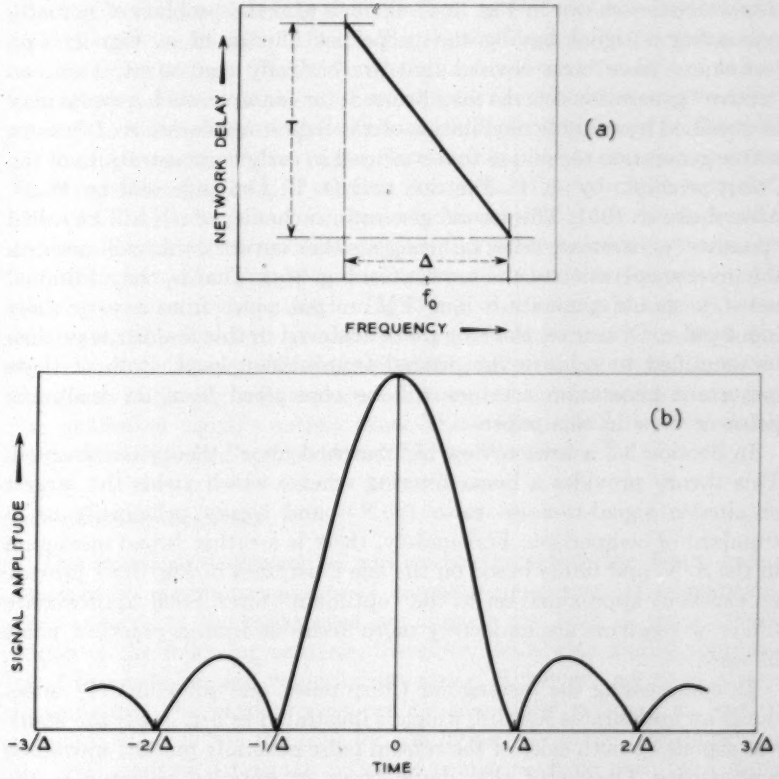


Fig. 3 — (a) Network delay vs. frequency characteristic suitable for phase equalization of the Chirp signal in Fig. 2; ideally, this network is chosen to have a flat loss characteristic. (b) Envelope of output response from the network in (a); this pulse now has a pulse width about $1/\Delta$ and an amplitude increase given by $\sqrt{D}$, where $D = T\Delta$ is called the dispersion factor.

illustrated in Fig. 3(b). In this expression, D, called the *dispersion factor*, is defined by the product $T\Delta$ and represents a parameter of fundamental significance. The collapsed pulse width, τ, is of the order of $1/\Delta$. Consequently, there has been a pulse width reduction given by the ratio

$$T/\tau \simeq T\Delta = D,$$

and a signal amplification given by $\sqrt{D}$. Useful systems already proposed will possess dispersion factors of 100, and even greater! Experimental measurements already made exhibit output signals such as Fig. 3(b) for $D = 120$.

Along with the problem of constructing delay networks having the characteristics shown in Fig. 3(a) there is also the problem of actually generating a signal having the properties illustrated in Fig. 2. Two techniques have been devised that are basically equivalent. First, an "active" generation scheme may be used; for example, such a signal may be obtained by suitable modulation of a voltage-tunable device. It was an active generation technique that was used in early demonstrations of the Chirp principle by A. F. Dietrich and O. E. DeLange and by W. J. Albersheim in 1951. The second generation scheme, which will be called "passive" generation, relies on using another network that will perform the inverse operation of the network in Fig. 3(a). That is, the additional network should generate a long FM output pulse from a very short input pulse. Of course, the long pulse achieved in this manner may then be amplified to achieve the desired transmission level. Both of these important generation schemes will be considered from an analytical point of view in this paper.

In Section 3.3 a brief review of "matched filter" theory is presented. This theory provides a linear filtering scheme which yields the largest obtainable signal-to-noise ratio (S/N), and serves principally as a standard of comparison. Fortunately, there is a rather broad maximum in the S/N, and filters based on the one illustrated in Fig. 3(a) provide an excellent approximation to the "optimum" filter. Such approximate filters or receivers are eminently more desirable from a practical point of view.

In compressing the rectangular Chirp pulse, one unfortunately introduces an undesirable feature, which is illustrated in Fig. 3(b): the spurious signals on both sides of the central pulse generally present unwanted ambiguities. Curves of this shape occur as antenna patterns in the theory of antenna design. By analogy, these spurious signals shall be referred to as *side lobes*. Techniques for reducing the relative side-lobe level have been extensively studied in antenna design. The analysis presented in Sections 3.3 and 3.4 to reduce the side lobes of the Chirp signal

will draw heavily on these antenna analogs. In particular, considerable side-lobe reduction may be obtained by properly attenuating the frequencies at the edges of the band Δ. Such a reduction is accompanied by a slight, but inevitable, increase in the pulse width.

The design of a practical radar system is not complete unless it is accompanied by a set of acceptable equipment tolerances; Section 3.4 discusses this important problem. The basis of the analytical study in this section is the *paired echo* concept introduced by MacColl.[8] Following a review of MacColl's treatment of the problem, there is a discussion of the effect on the collapsed pulse of phase and amplitude distortions. Also in Section 3.4 there appears a detailed discussion of spectrum shaping networks to reduce the side-lobe level based on Taylor's[9] approximation to the "ideal" Dolph-Tchebycheff signal. To present a clear and intuitive picture, Taylor's model is analyzed with the help of paired-echo theory. Effects on the pulse width and side-lobe level are presented, as well as a discussion of the influence of Taylor weighting on the S/N.

The final section of this paper, Section 3.5, is devoted to the unusual effects produced by targets moving in a radial direction with respect to the radar. As is well known, this relative motion gives rise to a small frequency shift of the returning signal; this phenomena is known as the *Doppler effect*. The principle qualitative effect of this frequency shift may be obtained by referring to Fig. 3(a), which shows the delay versus frequency characteristic of the receiver network. One sees from this figure that a uniform shift of the returning spectrum upwards (or downwards) in frequency will produce a collapsed pulse that will lead (or lag) the unaffected signal's output time. These effects will be discussed in terms of *ambiguity diagrams* introduced by Woodward.[10]

III. DETAILED ANALYSIS

The practical realization of any radar, including the Chirp radar, involves an enormous amount of complex equipment. It is gratifying that analysis of the radar need not include the bulk of these details. Much of the following analysis, therefore, deals with highly simplified ideal transmission and reception processes. However, one must always be sure that the ideal process can be essentially realized, at least in principle. This is not the case for the frequently made replacement of the true transmitted signal, which of course must be real, by a signal that is a complex function of time. The validity of considering complex signals is a simple consequence of the assumed linearity of the system. However, complex signals are useful, not only because the system is linear but also because they permit important simplifications regarding

the assumed receiving network characteristics to be made. In particular, the analysis in Section 3.1 will be concerned only with the positive frequency portion of signals and with networks having desired characteristics in this frequency range. The true signal output is obtained as usual by taking, say, the real part of the complex response.

The initial part of the analysis is devoted to studying the rectangular signal with linear FM. Crudely speaking, this may be regarded as assuming an "active" generation scheme; i.e., the problem of obtaining the signal illustrated in Fig. 2 is left to the engineer. However, this is not an accurate statement because, from an analytical point of view, consideration of the long rectangular signal in no way prejudges its means of generation. The discussion on the very important "passive" technique to generate the long-duration signal can then draw on the preceding analysis regarding the collapse process of a long FM pulse. Furthermore, there are two related reasons why an intrinsic study of the rectangular signal has special significance: (a) the rectangular envelope represents the most efficient use of transmitting tubes (as remarked in Section I), and (b) even the signals passively generated, which, for economical tube operation are again subjected to amplitude-limiting, seem to approach the characteristics of the ideal rectangular signal.

3.1 *Analytical Discussion of the Reception and Collapse of the Chirp Radar Signal*

Let $e_1(t)$ denote the analytical waveform of a single pulse received from an isolated, stationary point target. This real signal is chosen for convenience to be the real part of a complex waveform $\epsilon_1(t)$, where

$$\epsilon_1(t) = \text{rect}\left(\frac{t}{T}\right)e^{2\pi i(f_0 t + kt^2/2)}. \tag{1}$$

As customary, f_0 denotes some suitable carrier frequency. The function, rect z, introduced by Woodward, is defined by the following relations:

$$\begin{aligned} \text{rect } z &= 1, \quad \text{if } |z| < \tfrac{1}{2} \\ &= 0, \quad \text{if } |z| > \tfrac{1}{2}. \end{aligned} \tag{2}$$

Therefore, the received signal (1) is conveniently, but arbitarily, selected to be of unit amplitude. The particular phase characteristic in this signal is chosen to duplicate the linear FM properties illustrated in Fig. 2. With the phase of (1) represented by $\varphi = 2\pi(f_0 t + kt^2/2)$, the instantaneous signal frequency is defined, as usual, by

$$f_i = \frac{1}{2\pi}\frac{d\varphi}{dt} = f_0 + kt. \tag{3}$$

Thus, during the T-second interval of the pulse, the instantaneous frequency changes in a linear fashion from $f_0 - kT/2$ to $f_0 + kT/2$ [see Fig. 2(b)]. The net frequency sweep, Δ, is then the difference of these two values, or kT. This frequency interval Δ provides a convenient means to introduce natural units into the Chirp radar problem. The natural variables in which to measure time and frequency are expressed as follows:

$$y = t\Delta, \qquad \text{a "time" variable,} \tag{4a}$$

$$x = f/\Delta, \qquad \text{a "frequency" variable.} \tag{4b}$$

In conjunction with these natural variables the dimensionless product $T\Delta$ frequently appears. This product is called the *dispersion factor* and is denoted by D:

$$D \equiv T\Delta. \tag{5}$$

The complex received waveform (1) takes the following form when expressed in terms of the natural time and frequency variables:

$$\epsilon_1(y) = \text{rect}\left(\frac{y}{D}\right) e^{2\pi i(x_0 y + y^2/2D)}, \tag{6}$$

where $x_0 \equiv f_0/\Delta$ and $k = \Delta/T$ have been used. This signal is now D units of "time" in duration.

The signal illustrated in (1) or (6) possesses, by construction, the characteristics of Fig. 2. However, the qualitative argument regarding the instantaneous frequency does not necessarily provide an accurate description of the signal's frequency content. Such an accurate picture is obtained by studying the signal's Fourier transform. Let $\tilde{A}(f)$ denote the Fourier transform of an arbitrary function of time, $A(t)$.

The transform of the signal in (1) becomes

$$\begin{aligned} \tilde{\epsilon}_1(f) &= \int_{-\infty}^{\infty} \epsilon_1(t) e^{-2\pi i f t}\, dt \\ &= \int_{-T/2}^{T/2} e^{2\pi i[(f_0 - f)t + kt^2/2]}\, dt. \end{aligned} \tag{7a}$$

Following some algebraic manipulations, (7a) becomes

$$\tilde{\epsilon}_1(f) = \sqrt{\frac{T}{2\Delta}}\, e^{-i\pi(f - f_0)^2/k}[Z(u_2) - Z(u_1)], \tag{7b}$$

where $Z(u)$ is the complex Fresnel integral:

$$Z(u) = C(u) + iS(u) = \int_0^u e^{i\pi\alpha^2/2}\, d\alpha. \tag{8}$$

The arguments u_2 and u_1 are defined by

$$u_2 = -2(f - f_0)\sqrt{\frac{T}{2\Delta}} + \sqrt{\frac{T\Delta}{2}}, \tag{9a}$$

$$u_1 = -2(f - f_0)\sqrt{\frac{T}{2\Delta}} - \sqrt{\frac{T\Delta}{2}}. \tag{9b}$$

The shape of the frequency spectrum, as would be determined, for example, by a spectrum analyser, contains only the amplitude information. The amplitude of the signal spectrum in (7) is obtained simply by taking the absolute value:

$$\begin{aligned} |\bar{\epsilon}_1(f)| &= \sqrt{\frac{T}{2\Delta}}\,|Z(u_2) - Z(u_1)| \\ &= \sqrt{\frac{T}{2\Delta}}\,\{[C(u_2) - C(u_1)]^2 + [S(u_2) - S(u_1)]^2\}^{\frac{1}{2}}, \end{aligned} \tag{10a}$$

where $C(u)$ and $S(u)$ are defined in (8) as the real and imaginary part of $Z(u)$. In establishing the relative frequency content, a normalized form of (10a) is used. Figs. 4, 5 and 6 illustrate various spectra obtained by plotting

$$\frac{1}{\sqrt{2}}\,\{[C(u_2) - C(u_1)]^2 + [S(u_2) - S(u_1)]^2\}^{\frac{1}{2}}, \tag{10b}$$

where u_2 and u_1 appear in (9a) and (9b). Inspection of (9a) and (9b)

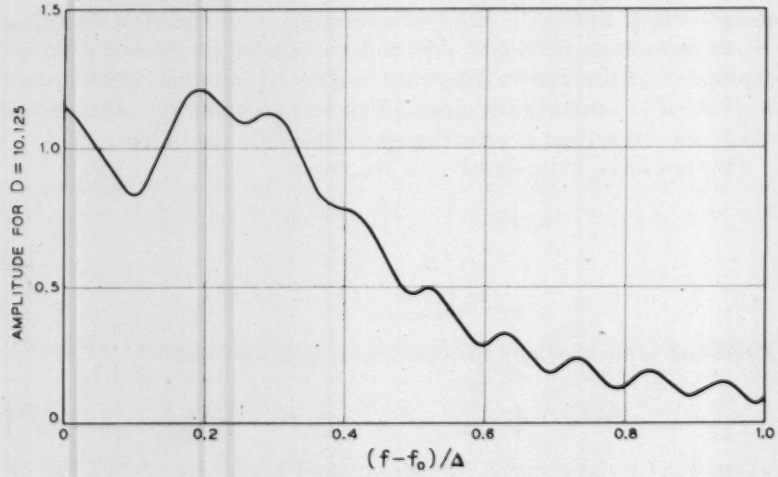


Fig. 4 — Spectral amplitude of a rectangular Chirp signal for $D = 10.125$. The shape is symmetric about the point $(f - f_0)/\Delta = 0$.

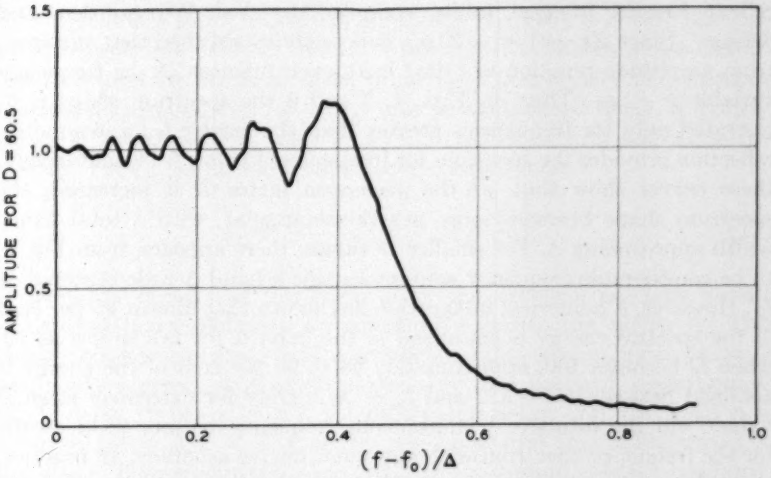


Fig. 5 — Spectral amplitude of a rectangular Chirp signal for $D = 60.5$.

shows that the dispersion factor, D, is the only auxiliary variable present besides the "frequency" $x - x_0$, i.e. $(f - f_0)/\Delta$. The values of D chosen for the curves in Figs. 4, 5 and 6 were selected so that $\sqrt{2D}$ would be rational. In this manner full use (no interpolation) was made of the ex-

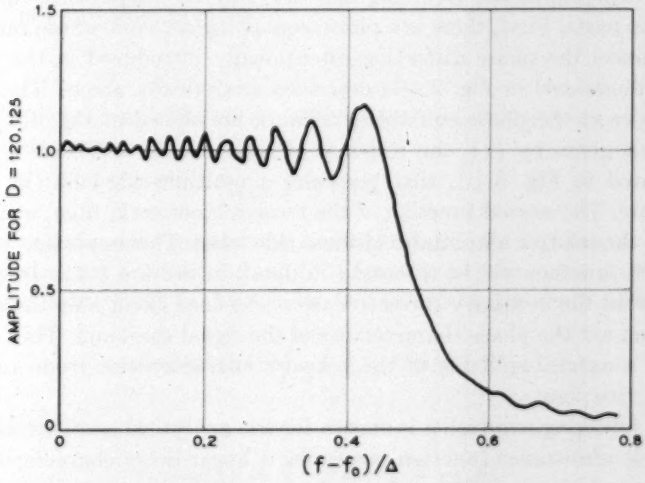


Fig. 6 — Spectral amplitude of a rectangular Chirp signal for $D = 120.125$.

cellent Fresnel integral tables compiled by Van Wijngaarden and Scheen.[11] Since $Z(-u) = -Z(u)$, it is easily established that the spectrum amplitude function in (10b) is an even function of the frequency variable $x - x_0$. Thus, in Figs. 4, 5 and 6 the spectrum shape is illustrated only for frequencies greater than the carrier f_0; a symmetric reflection provides the spectrum for frequencies f below f_0. Qualitatively, these curves show that, as the dispersion factor D is increased, the spectrum shape becomes more nearly rectangular, with a total bandwidth approaching Δ. For smaller D values, there appears, from Fig. 4, to be considerable frequency content outside a band Δ wide centered at f_0. However, a numerical integration has shown that almost 95 per cent of the spectral energy is contained in the band Δ for D's as low as 10; when D becomes 100, approximately 98 to 99 per cent of the energy is confined between $f_0 - \Delta/2$ and $f_0 + \Delta/2$. Only for extremely large D values will the intuitive instantaneous frequency picture yield results for the frequency spectrum with any quantitative accuracy. In practice, the value of f_0, whether at radio frequency (RF) or some intermediate frequency (IF), will always be at least several times the approximate bandwidth Δ. So long as f_0 continues to be several times Δ, Fourier analysis shows that the complex radar signal in (1) contains only positive frequencies to a very high degree of accuracy. As stated previously, simplified receiver network characteristics may be assumed when the ideal received waveform contains only positive frequencies.

The function of the receiving network may be conveniently divided into two parts. First, there is a phase-equalizing network whose function is to cancel the phase distortion intentionally introduced in the Chirp signal illustrated in Fig. 2 and described analytically above. The characteristics of the phase-equalizing network are shown in Fig. 3(a). For an input given by (1), the response of this network will have a form, illustrated in Fig. 3(b), that possesses a multiple-side-lobe (in time) structure. The second function of the receiving network, then, will be to reduce the relative amplitudes of these side lobes. This important second network function will be discussed in detail in Section 3.4.2. It will be found that the necessary corrective networks need affect only the amplitude and not the phase characteristic of the signal spectrum. Thus, there is also a natural splitting of the network characteristics from an analytical viewpoint.

The following expression is chosen for the analytical specification of a network admittance function possessing a linear delay characteristic:

$$\bar{Y}(f) = e^{i\pi p(f-f_0)^2}. \tag{11}$$

This network is given a uniform amplitude characteristic with the particular amplitude value of unity. To see that the chosen phase characteristic represents the linear delay property, it is necessary to investigate the instantaneous delay of $\tilde{Y}(f)$. The instantaneous delay is proportional to the phase $\psi = \pi p(f - f_0)^2$ differentiated with respect to frequency. Let t_d denote the instantaneous delay; then

$$t_d \equiv -\frac{1}{2\pi}\frac{d\psi}{df} = -p(f - f_0). \tag{12}$$

It should be noted that this linear delay characteristic has been chosen so as to give zero delay to a frequency f equal to the carrier f_0. This choice is made for analytical convenience only, and a linear phase is frequently ignored, since it gives rise only to a uniform, distortionless signal delay. When $p = k^{-1}$, the differential delay over the band Δ is given by $p\Delta = \Delta/k = T$. It is this particular case that is illustrated in Fig. 3(a). Throughout this paper only this optimum case is considered:

$$\tilde{Y}(f) = e^{i\pi(f-f_0)^2/k}. \tag{13}$$

The detailed response of the lossless phase equalizer to the input Chirp signal, (1), will now be investigated. The complex response function will be denoted by $\epsilon(t)$ and its Fourier transform by $\tilde{\epsilon}(f)$. The true, real response $e(t)$ is obtained, as usual, by taking the real part of $\epsilon(t)$: $e(t) = \text{Re}\ \epsilon(t)$. From the definition of the network admittance function, the output spectrum is given by

$$\tilde{\epsilon}(f) = \tilde{Y}(f)\tilde{\epsilon}_1(f). \tag{14}$$

A well-known result of Fourier analysis permits the response $\epsilon(t)$ to be defined according to

$$\epsilon(t) = Y(t) * \epsilon_1(t). \tag{15}$$

Here $Y(t)$, the network impulse response, is simply the Fourier transform of the network function in (13). The star $(*)$ denotes the convolution operation, which is defined for two arbitrary functions as follows:

$$f(z) * g(z) = \int_{-\infty}^{\infty} f(z - w)g(w)\, dw, \tag{16a}$$

or equivalently

$$f(z) * g(z) = \int_{-\infty}^{\infty} f(v)g(z - v)\, dv. \tag{16b}$$

It is somewhat easier to calculate the response $\epsilon(t)$ by using (15). First, the impulse response, $Y(t)$, must be computed:

$$\begin{aligned} Y(t) &= \int_{-\infty}^{\infty} \tilde{Y}(f) e^{2\pi i f t}\, df \\ &= \int_{-\infty}^{\infty} e^{2\pi i [f t + (f-f_0)^2/2k]}\, df. \end{aligned} \tag{17}$$

This is a well-known integral,† and (17) is evaluated as

$$Y(t) = \sqrt{\frac{i\Delta}{T}}\, e^{2\pi i (f_0 t - k t^2/2)}. \tag{18}$$

The complex output, $\epsilon(t)$, is obtained by forming the convolution product of $Y(t)$ and $\epsilon_1(t)$ given, respectively, in (18) and (1). Thus,

$$\epsilon(t) = \sqrt{\frac{i\Delta}{T}} \int_{T/2}^{T/2} e^{2\pi i [f_0 t + k\tau^2/2 - k(t-\tau)^2/2]}\, d\tau,$$

which simplifies considerably, due to cancellations in the exponent, to

$$\epsilon(t) = \sqrt{\frac{i\Delta}{T}}\, e^{2\pi i (f_0 t - k t^2/2)} \int_{-T/2}^{T/2} e^{2\pi i k t \tau}\, d\tau.$$

Carrying out the elementary integral that remains,

$$\epsilon(t) = \frac{1}{\pi k t} \sqrt{\frac{i\Delta}{T}} \sin(\pi k t T) e^{2\pi i (f_0 t - k t^2/2)},$$

which, when regrouped becomes

$$\epsilon(t) = \sqrt{Di}\, \frac{\sin \pi \Delta t}{\pi \Delta t}\, e^{2\pi i (f_0 t - k t^2/2)}. \tag{19}$$

This response assumes the following form when expressed in terms of the natural time variable $y = \Delta t$:

$$\epsilon(y) = \sqrt{D}\, \frac{\sin \pi y}{\pi y}\, e^{2\pi i (x_0 y - y^2/2D) + i\pi/4}. \tag{20}$$

Following Woodward's nomenclature[10] it is useful to introduce

$$\operatorname{sinc} y \equiv \frac{\sin \pi y}{\pi y} \tag{21}$$

as a convenient shorthand. The functions sinc y and rect x are Fourier transforms of one another. The envelope of the complex response is given by the absolute value of (19) and takes the form

† This transform is number 708.0 in Campbell and Foster.[12]

$$\sqrt{D}\left|\frac{\sin \pi\Delta t}{\pi\Delta t}\right| = \sqrt{D}\,|\,\mathrm{sinc}\,\Delta t\,|\,. \tag{22}$$

It is this envelope that is shown in Fig. 3(b): this collapsed pulse has an amplitude increase of $\sqrt{D}$ over the input pulse and a new pulse width $\tau \sim 1/\Delta$, measured at the half-power points (3-db points).

There are two important facts to be noted regarding the complex response signal given in (19) and (20). First, there exists a residual FM after phase equalization. By comparing (19) with the input signal $\epsilon_1(t)$ in (1) one easily sees that the residual FM occurs at the same rate, but is reversed in sense. Otherwise stated, if $k > 0$, then the frequency of the transmitted pulse *rises* during the pulse period, while the instantaneous frequency of the collapsed signal (19) *falls*, but at the same absolute rate. This effect has been observed in a laboratory model for a low dispersion factor ($D \approx 10$). One would expect to observe this effect principally for low D's, as may be seen from the following qualitative argument. Using (20), the instantaneous "frequency" is given by $x_0 - y/D$. Now, about 90 per cent of the signal energy is confined in the "time" interval $|\,y\,| < 1$ by the envelope sinc y. If attention is confined to this region of pronounced signal amplitude, the frequency will vary only between $x_0 + 1/D$ and $x_0 - 1/D$. From the discussion following (10b), the value of $x_0 = f_0/\Delta$ will be seen to lie generally in the range $x_0 > 2$. For large D, the frequency in the significant region of the output signal remains constant to an accuracy of about $1/Dx_0$. As D becomes larger, deviations from a constant frequency are therefore nearly undetectable. A slight reservation to this point will be noted in Section 3.3 but, for the most part (especially in studying the side-lobe reduction in Section 3.4.2), the residual FM in the response will be ignored.

The second important point to be noted regarding the output signal (19) pertains to the envelope, sinc Δt. Although the initial envelope, rect (t/T), and the final one, sinc Δt, are both functions of time, there does appear to be a Fourier transform relation between the *functional* form of these envelopes. The two points just stressed above will now be shown to be consequences of an important result of greater generality.

Assume for the moment that, instead of (1), the following complex signal is transmitted:

$$\epsilon'(t) = E(t)e^{2\pi i(f_0 t + kt^2/2)}. \tag{23}$$

This signal differs from (1) insofar as the envelope rect (t/T) has been replaced by an *arbitrary* envelope denoted by $E(t)$. The linear FM characteristic remains unchanged. Attention is now turned to finding the

response of the same lossless network given before in (13) to the modified input, $\epsilon'(t)$. The spectrum of the new output signal is given by

$$\begin{aligned} \tilde{\epsilon}''(f) &= \tilde{Y}(f)\tilde{\epsilon}'(f) \\ &= \int_{-\infty}^{\infty} E(\tau)e^{2\pi i[(f_0-f)\tau+k\tau^2/2+(f-f_0)^2/2k]}\,d\tau. \end{aligned} \tag{24a}$$

The terms in the brackets constitute a perfect square, so that

$$\tilde{\epsilon}''(f) = \int_{-\infty}^{\infty} E(\tau)e^{(\pi i/k)[(f-f_0)-k\tau]^2}\,d\tau. \tag{24b}$$

By applying another Fourier transform, the output signal $\epsilon''(t)$ is generated:

$$\epsilon''(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(\tau)e^{2\pi ift+\pi i/k[(f-f_0)-k\tau]^2}\,df\,d\tau. \tag{25}$$

The integration over f is similar to one carried out previously, and (25) becomes

$$\epsilon''(t) = \sqrt{ki}\,e^{2\pi i(f_0t-kt^2/2)}\int_{-\infty}^{\infty} E(\tau)e^{2\pi ikt\tau}\,d\tau. \tag{26}$$

The remaining integral in (26) essentially defines the Fourier transform of $E(t)$, so that our final expression for $\epsilon''(t)$ becomes

$$\epsilon''(t) = \sqrt{ki}\,\tilde{E}(-kt)e^{2\pi i(f_0t-kt^2/2)}. \tag{27}$$

This result shows that the two points stressed above are special cases of an exceedingly general result. That is, the output response in (27) possesses the same linear FM property as the input signal but with the reversed sense, and the Fourier-transform functional relation between input and output envelopes is established as a general rule.

The preceding general analysis for the arbitrary envelope $E(t)$ provides an excellent intuitive picture for the next topic to be discussed: the *generation* of the long transmitted pulse by passive means.

3.2 *Passive Generation of the Long FM Transmitter Signal*

In any scheme of generating the long-duration FM signal, the foremost objective is to secure a signal whose properties are similar to those illustrated in Fig. 2. The passive generation means described in this section can be no exception to this rule.

The previous section concluded with a general theorem relating the input and output envelopes of a signal with linear FM when passed

through a lossless, linear delay equalizer. This theorem illustrates how an input signal with a long-duration envelope becomes transformed by the equalizer to a signal whose new envelope is much shorter in duration. In addition, implicitly contained in this theorem is the effect produced by an input signal that possesses a very *short*-duration envelope. Since the output and input envelopes are functionally connected through the Fourier transform, the response envelope will be one of long-duration when a narrow input signal is used. Inspection of (23) and (27) shows that, if the initial envelope $E(t)$ is chosen as sinc Δt, then the envelope of the equalizer output is given by $\sqrt{k}\,\bar{E}(-kt) = \sqrt{k}\,\Delta^{-1}$ rect $(kt/\Delta) = D^{-\frac{1}{2}}$ rect (t/T). The full signal, with its linear FM characteristic, may then be amplified to achieve the proper transmission level.

There are two points that arise in connection with the passive generation scheme described above. The first deals with the particular characteristics of the narrow-input signal whose envelope $E(t) =$ sinc Δt. One of the additional requirements for this signal, according to (23), is a linear FM characteristic. In the discussion following (22) it was shown, for all cases of practical instance, that the linear FM characteristic would cause only a very slight deviation from a constant frequency. Qualitatively, it appears that the properties of the output signal (27) will remain essentially unchanged if the input signal contains *no* FM; i.e., if (23) is replaced by $\epsilon'(t) =$ (sinc $\Delta t)e^{2\pi i f_0 t}$. Practically, it is of course much easier to generate simply the required short envelope than a signal with such delicate FM properties. It is the neglect of the linear FM characteristic that is responsible for what analytical distinctions there are in studying the passive generation scheme.

The second issue to be discussed relates to the particular sign of the linear FM slope in the network output signal in (27). By comparing (27) and (1) it is clear that the response signal's FM is directed oppositely to the FM of the desired input in (1). This means that the same network *characteristics* cannot be used both to generate the long FM signal in the transmitter *and* to collapse that signal in the receiver. What is clearly needed, then, is two distinct network characteristics, one with $k > 0$ and the other with $k < 0$, so that the sum of the delays in these two characteristics is a constant, independent of frequency. If one of these characteristics is used in generating the long FM signal, then the remaining one represents the appropriate collapsing characteristic. The reciprocal relation existing between these two network characteristics is illustrated in Fig. 7.

In discussing the generation of the long FM signal by passive means it was stated that, for practical reasons, a short, constant-frequency

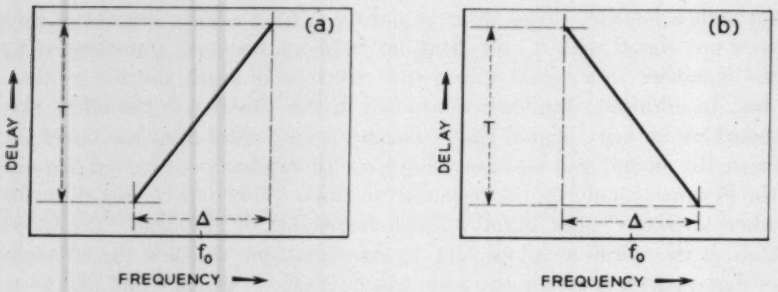


Fig. 7 — (a) Equivalent network characteristics for the transmitter in the passive generation scheme; a uniform amplitude characteristic is assumed along with the phase characteristic illustrated. (b) Appropriate receiver network characteristic to be used with the passive transmitter network in (a); note that the sum of the delay introduced by the transmitter and receiver networks is a constant.

input signal is used whose envelope is given by sinc Δt. The network output envelope will no longer match that shown in Fig. 2, due to the neglect of a small-input FM characteristic. It is important to determine the quantitative effect that this approximation has on the network output, since this is the characteristic signal transmitted by the radar. The equivalent transmitter network characteristics are chosen in accord with Fig. 7(a):

$$\bar{Y}'(f) = e^{-i\pi(f-f_0)^2/k}. \tag{28}$$

The complex input to this network is denoted by $\epsilon_2(t)$, where

$$\epsilon_2(t) = (\text{sinc } \Delta t)e^{2\pi i f_0 t}, \tag{29}$$

and again the shorthand sinc y has been used. The response of the transmitter network is given by

$$\begin{aligned} \epsilon_2'(t) &= \int_{-\infty}^{\infty} e^{2\pi i f t}\bar{\epsilon}_2(f)\bar{Y}'(f)\,df \\ &= \frac{1}{\Delta}\int_{f_0-\Delta/2}^{f_0+\Delta/2} e^{2\pi i[ft-(f-f_0)^2/2k]}\,df. \end{aligned} \tag{30}$$

This result, like (7), may be expressed in terms of the complex Fresnel integral $Z(u)$ defined in (8):

$$\epsilon_2'(t) = \frac{1}{\sqrt{2\Delta T}} e^{2\pi i(f_0 t + kt^2/2)}[Z^*(v_2) - Z^*(v_1)], \tag{31}$$

where the superscript star (*) denotes complex conjugate. The arguments v_2 and v_1 are given by

$$v_2 = \sqrt{2}\left(\sqrt{\frac{\Delta}{T}}\, t + \frac{1}{2}\sqrt{\Delta T}\right), \tag{32a}$$

$$v_1 = \sqrt{2}\left(\sqrt{\frac{\Delta}{T}}\, t - \frac{1}{2}\sqrt{\Delta T}\right). \tag{32b}$$

What is the relation between the desired rectangular envelope and the true one given in (31)? Apart from a scale factor, this envelope is given by

$$\frac{1}{\sqrt{2}}\{[C(v_2) - C(v_1)]^2 + [S(v_2) - S(v_1)]^2\}^{\frac{1}{2}}. \tag{33}$$

Notice that this envelope is of precisely the same functional form as (10b), which describes the amplitude of the frequency spectrum for the rectangular Chirp signal in (1). Consequently, by merely reinterpreting Figs. 4, 5 and 6, curves representing the transmitted signal envelope may be obtained. To facilitate the reinterpretation the definitions of the arguments used in calculating (10b) are repeated here [see (9a) and (9b)]:

$$u_2 = \sqrt{\frac{2}{D}}\left[-D(x - x_0) + \frac{D}{2}\right],$$

$$u_1 = \sqrt{\frac{2}{D}}\left[-D(x - x_0) - \frac{D}{2}\right].$$

A comparison with the new arguments in (32) for the transmitted envelope shows that the appropriate reinterpretation of the abscissa is given by the transformation: $(x - x_0) \rightarrow y/D = t/T$. Thus, Fig. 4, for example, shows one-half of the symmetric envelope for the case $D = 10.125$; the abscissa value of 0.5 is to be understood as the value of the parameter t/T. This value of t/T marks the boundary in time of the desired rectangular envelope. The value 1.0 on the ordinate represents the corresponding amplitude of the ideal rectangular envelope. A similar abscissa relabeling and reinterpretation applies to Figs. 5 and 6. These latter two figures illustrate that, as D becomes larger, the transmitted signal envelope approaches the rectangular envelope more closely. The deviation from a rectangular shape is entirely due to the neglect of the small linear FM in the input.

After the transmitted signal is reflected from, say, a point target, the return signal undergoes phase equalization in the receiver. Although Section 3.1 was devoted to the properties of a collapsed pulse, it is instructive to see what new shape is obtained when the transmitted envelope is no longer exactly rectangular. The calculation of the final

response signal is almost immediate. The short signal in the transmitter is given in (29): $\epsilon_2(t) = (\text{sinc } \Delta t)e^{2\pi i f_0 t}$. The transmitter network that spreads out this short pulse has a characteristic given in (28): $\tilde{Y}'(f) = e^{-i\pi(f-f_0)^2/k}$. The receiver network characteristic to collapse the transmitted signal appears in (13): $\tilde{Y}(f) = e^{i\pi(f-f_0)^2/k}$. The ideal receiver output, then, is simply

$$\begin{aligned} \epsilon_2''(t) &= \int_{-\infty}^{\infty} \tilde{Y}(f)\tilde{Y}'(f)\tilde{\epsilon}_2(f)e^{2\pi i f t}\, df \\ &= \int_{-\infty}^{\infty} \tilde{\epsilon}_2(f)e^{2\pi i f t}\, df \\ &= \epsilon_2(t). \end{aligned} \tag{34}$$

That is, the receiver equalizer output equals the short generating signal employed in the transmitter: $\epsilon_2''(t) = (\text{sinc } \Delta t)e^{2\pi i f_0 t}$. It is to be noted that the *envelope* of the response, sinc Δt, achieved by passive generation is equivalent to the *envelope* obtained from an "active" generation given in (19) and (22). The only distinction in the two modes of generation rests in the presence or absence of a small, residual linear FM. One possible consequence of this small distinction will be discussed in the next section.

3.3 *Some Signal-to-Noise Considerations in the Study of Chirp Radars*

Up to this point the analysis has been confined only to uniformly lossless delay-equalizing networks. In practice, of course, this idealization cannot hold true, nor would one even want it to hold true. For a study of the S/N properties of Chirp radars, additional networks must be used which shape the spectrum by introducing some loss. Spectrum shaping will be discussed again in Section 3.4.2 from a different point of view.

3.3.1 *Maximum Signal-to-Noise and the Matched Filter*

Since the presence of noise represents a degradation of signal information, it is important that the ratio of signal to noise be made as large as practical. Woodward[10] discusses several different criteria for just which quantity should be maximized. For this paper, the peak signal power to mean noise power ratio is selected as the definition of S/N. The noise perturbing the ideal Chirp system is assumed to be additive and to have the properties of so-called white gaussian noise. Such noise is (a) statistically independent at each frequency, (b) of uniform mean power density independent of frequency and (c) gaussian in its amplitude dis-

tribution at each frequency. With no loss of generality, the mean noise power can be normalized to take the value one. Therefore, the total noise output is given by the noise energy in a "typical" output noise signal whose Fourier transform is denoted by $\tilde{Y}(f)\tilde{n}(f)$, where $\tilde{Y}(f)$ denotes an arbitrary network characteristic. Using the criteria for mean noise level, the noise power output becomes

$$N \equiv \int |\tilde{Y}(f)|^2 |\tilde{n}(f)|^2 df = \int |\tilde{Y}(f)|^2 df, \tag{35}$$

when expressed in terms of the arbitrary filter $\tilde{Y}(f)$. It is clear that additional spectrum shaping of the linear delay network is necessary to give meaning to the noise defined by (35).

In order to make signal-to-noise comparisons meaningful, both the signals and the noise must be normalized. The common practice is to normalize the signal power by dividing it by the total energy content in the signal: $\int |\epsilon(t)|^2 dt = \int |\tilde{\epsilon}(f)|^2 df$, where $\epsilon(t)$ now denotes an arbitrary signal chosen, for convenience, to have only positive frequencies. The normalized peak signal power, S, is then defined by

$$S = (\text{maximum over } t) \frac{\left| \int \tilde{Y}(f)\tilde{\epsilon}(f)e^{2\pi i f t} df \right|^2}{\int |\tilde{\epsilon}(f)|^2 df}. \tag{36}$$

Finally, the S/N is given by

$$S/N = \frac{\left| \int \tilde{Y}(f)\tilde{\epsilon}(f)e^{2\pi i f t_M} df \right|^2}{\int |\tilde{\epsilon}(f)|^2 df \int |\tilde{Y}(f)|^2 df}, \tag{37}$$

where the maximum of the right-hand side occurs at $t = t_M$; t_M will in general be a function of both $\tilde{\epsilon}(f)$ and $\tilde{Y}(f)$. Suppose now that $\tilde{\epsilon}(f)$ is fixed. For which filter $\tilde{Y}(f)$ will the value of S/N achieve its maximum value? This is a problem in the calculus of variations whose solution, for present purposes, is summarized in the well-known Schwarz inequality, which states, for two arbitrary functions $f(z)$ and $g(z)$, that

$$\frac{\left| \int f(z)g(z)\, dz \right|^2}{\int |f(v)|^2 dv \int |g(w)|^2 dw} \leqq 1. \tag{38}$$

The equality holds only when $f(z)$ is proportional to $g^*(z)$. Upon applying this theorem to (37), the maximum S/N is achieved when $\tilde{Y}(f)$ is proportional to $\tilde{\epsilon}^*(f)e^{-2\pi i f t_M}$. The exponential factor merely represents a gross delay without distortion, which will be ignored. Consequently, the *matched filter*, $\tilde{Y}_m(f)$, for the signal $\epsilon(t)$ is defined as

$$\tilde{Y}_m(f) \equiv \tilde{\epsilon}^*(f). \tag{39}$$

It should be noticed that the particular value of the maximum S/N is in no way related to which specific signal and matched filter are used; any signal coupled with its matched filter attains the same maximum value of "one".

It is important to find the specification of the output response when a matched filter is used. According to (39), the complex response $\epsilon_m(t)$ is determined by

$$\epsilon_m(t) = \int_{-\infty}^{\infty} |\epsilon(f)|^2 e^{2\pi i f t}\, df, \tag{40a}$$

which may be re-expressed with the aid of the convolution operator defined in (16):

$$\begin{aligned} \epsilon_m(t) &= \epsilon^*(-t) * \epsilon(t) \\ &= \int_{-\infty}^{\infty} \epsilon^*(\tau - t)\epsilon(\tau)\, d\tau. \end{aligned} \tag{40b}$$

It follows from (40b) that the complex response from a matched filter is determined by a correlation process. As usual, the true, real response is obtained by taking the real part of $\epsilon_m(t)$ in (40a) or (40b). The matched filter for Chirp radars can now be studied with the aid of the general theory just presented.

3.3.2 *Matched Filter for Chirp*

3.3.2.1 *Matched Filter for the Active Generation Case.* For the case of active generation, the initial signal is $\epsilon_1(t)$ in (1), namely,

$$\epsilon_1(t) = \text{rect}\left(\frac{t}{T}\right) e^{2\pi i(f_0 t + k t^2/2)},$$

where rect z is defined by (2). The matched filter for this signal, $\tilde{Y}_{m_1}(f)$, may be readily obtained from the analysis of the spectrum of $\epsilon_1(t)$ in (7):

$$\begin{aligned} \tilde{Y}_{m_1}(f) &= \tilde{\epsilon}_1^*(f) \\ &= \sqrt{\frac{T}{2\Delta}}\, e^{+i\pi(f-f_0)^2/k}[Z^*(u_2) - Z^*(u_1)], \end{aligned} \tag{41}$$

where u_2 and u_1 are given in (9). The matched filter impulse response is also readily obtained by using the general relation: $Y_m(t) = \epsilon^*(-t)$. Therefore,

$$\begin{aligned} Y_{m_1}(t) &= \epsilon_1^*(-t) \\ &= \text{rect}\left(\frac{t}{T}\right) e^{2\pi i(f_0 t - kt^2/2)}. \end{aligned} \tag{42}$$

The output response when the rectangular Chirp signal is passed through its own matched filter may be computed with the aid of (40b):

$$\epsilon_{m_1}(t) = \int_{-\infty}^{\infty} \text{rect}\left(\frac{\tau - t}{T}\right) \text{rect}\left(\frac{\tau}{T}\right) e^{2\pi i[f_0 t + (k/2)\tau^2 - (k/2)(\tau - t)^2]} d\tau.$$

When $0 \leqq t \leqq T$,

$$\epsilon_{m_1}(t) = e^{2\pi i(f_0 t - kt^2/2)} \int_{t-T/2}^{T/2} e^{2\pi i k\tau t}\, d\tau.$$

Finally, after the remaining integration is performed,

$$\epsilon_{m_1}(t) = \frac{1}{\pi kt} e^{2\pi i f_0 t} \sin \pi\,(ktT - kt^2). \tag{43}$$

It is readily determined that the envelope of $\epsilon_{m_1}(t)$ is an even function of time. Therefore,

$$T \frac{\sin \pi(\Delta \mid t \mid - kt^2)}{\pi\Delta \mid t \mid} \tag{44a}$$

represents the envelope of the matched filter response for times $\mid t \mid \leqq T$. For times $\mid t \mid > T$ the envelope vanishes. When (44a) is expressed in terms of the natural time variable, $y = \Delta t$, the envelope becomes

$$\sqrt{D} \frac{\sin \pi\left(\mid y \mid - \dfrac{y^2}{D}\right)}{\pi \mid y \mid}, \tag{44b}$$

where a normalization has been chosen so as to facilitate a direct comparison with the response envelope in (22) obtained by using the lossless phase equalizer [see (13)]. The output from the simple linear delay equalizer, $\sqrt{D}(\sin \pi y)/\pi y$, is a remarkable approximation to the response from the more complicated matched filter, $\tilde{Y}_{m_1}(f)$, especially as the dispersion factor increases. The similarity between the two signals does not mean that they possess similar values of S/N, for, as already noted, the noise energy passed by the ideal linear delay equalizer is infinite. The signal similarity does suggest, however, that a large S/N

value may be obtained by combining with the delay equalizer a shaping network that limits the noise but does not seriously attenuate the strong signal frequencies.

As an extreme example of such a shaping network, the S/N properties will be investigated when an ideal filter, which suppresses all frequencies outside the range $|f - f_0| < \Delta/2$, is added to the delay equalizer. The combined network is specified by

$$\tilde{Y}_{rect}(f) = \text{rect}\left(\frac{f - f_0}{\Delta}\right) e^{i\pi(f-f_0)^2/k}. \tag{45}$$

The total noise passed through $\tilde{Y}_{rect}$ may be computed from the general definition in (35):

$$N_{rect} = \int |\tilde{Y}_{rect}(f)|^2 \, df = \Delta. \tag{46}$$

From symmetry arguments, the peak output signal value should occur at $t = 0$. The response at $t = 0$ is given by

$$\begin{aligned} \epsilon_{rect}(0) &= \int \tilde{Y}_{rect}(f)\tilde{\epsilon}_1(f)\, df \\ &= \sqrt{\frac{T}{2\Delta}} \int_{f_0-\Delta/2}^{f_0+\Delta/2} [Z(u_2) - Z(u_1)]\, df, \end{aligned} \tag{47a}$$

where u_2 and u_1 are linear functions of f. If natural variables are introduced and a shift of the origin is made, (47a) is transformed into

$$\begin{aligned} \epsilon_{rect}(0) = \sqrt{\frac{D}{2}} \int_{-1}^{1} \Bigg\{ & Z\left[\sqrt{\frac{D}{2}}\,(2x+1)\right] \\ & - Z\left[\sqrt{\frac{D}{2}}\,(2x-1)\right]\Bigg\} dx, \end{aligned} \tag{47b}$$

where the definition of u_2 and u_1 is equivalent to that given in (9). The integrals appearing in (47b) may be evaluated with the aid of the following general formula:

$$\begin{aligned} \int_a^b Z(\alpha x + \beta)\, dx = \frac{1}{\alpha}\,[(\alpha b + \beta)Z(\alpha b + \beta) - (\alpha a + \beta)Z(\alpha a + \beta)] \\ + \frac{i}{\pi\alpha}\,[e^{i\pi(\alpha b+\beta)^2/2} - e^{i\pi(\alpha a+\beta)^2/2}]. \end{aligned}$$

If this general formula is applied to the specific calculation required in (47b), it follows that

$$\epsilon_{\text{rect}}(0) = \sqrt{2D}\,Z(\sqrt{2D}) - \frac{i}{\pi}(1 - e^{i\pi D}). \tag{48}$$

It is simpler, and sufficient for present purposes, to confine attention to D values that are even integers. With this restriction, (48) reduces to

$$\epsilon_{\text{rect}}(0) = \sqrt{2D}Z(\sqrt{2D}),$$

and the peak signal power is given by

$$|\,\epsilon_{\text{rect}}(0)\,|^2 = 2D\,|\,Z(\sqrt{2D})\,|^2. \tag{49}$$

To complete the calculation for the normalized peak signal power, S, it is necessary, according to (36), to divide (49) by $\int |\,\epsilon_1(t)\,|^2\,dt = \int \text{rect}\,(t/T)\,dt = T$. Finally, the S/N is

$$(\text{S/N})_{\text{rect}} = 2\,|\,Z(\sqrt{2D})\,|^2, \tag{50}$$

where $Z(u)$ is the complex Fresnel integral defined in (8). Any decrease of S/N below the matched filter maximum value of "one" represents a S/N degradation. In Fig. 8, the S/N degradation in decibels [i.e. units of 10 $\log_{10}$ (S/N)] is illustrated for various values of D, the dispersion factor. Only even integral values of D were used in calculating this result; a smooth curve was then drawn between the calculated points. Certainly no qualitative error is made in this process. As anticipated, Fig. 8 shows that, as D becomes larger, the simple filter characteristics specified by $\tilde{Y}_{\text{rect}}$ approach those of the matched filter. The extremely small S/N degradation, especially for large D, shows a clear practical preference for characteristics like those of $\tilde{Y}_{\text{rect}}$ as compared to, say, the complicated amplitude characteristic required in building a matched filter. According to (39), the matched filter amplitude characteristic must match the signal spectrum and, for example, would need to resemble one of the spectra shown in Figs. 4, 5 or 6.

3.3.2.2 *Matched Filter for the Passive Generation Case.* Suppose, instead of the rectangular envelope, that the transmitted waveform had an envelope characteristic of the *passive* generation scheme discussed in Section 3.2. Under ideal circumstances a short signal, $(\text{sinc}\,\Delta t)e^{2\pi i f_0 t}$, is dispersed in the transmitter by a suitable linear delay equalizer. The resulting transmitted signal has a spectrum proportional to

$$\text{rect}\left(\frac{f - f_0}{\Delta}\right) e^{-i\pi(f-f_0)^2/k}.$$

From the definition of a matched filter in (39), the filter $\tilde{Y}_{\text{rect}}(f)$ in (45)

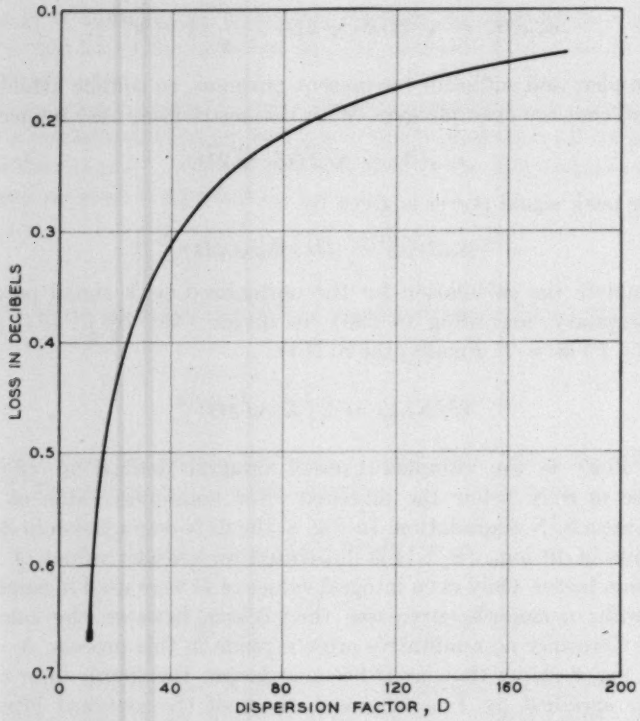


Fig. 8 — Degradation in S/N from ideal maximum when the rectangular Chirp signal passes through a delay equalizer and a sharp cutoff filter of width Δ. The loss decreases rapidly as the dispersion factor, D, increases.

actually represents the matched filter in the case of ideal passive generation. With this filter the response becomes simply $(\text{sinc } \Delta t)e^{2\pi i f_0 t}$, as was noted in Section 3.2; now it appears that this particular network also yields the maximum S/N value of "one".

3.3.3 *Signal-to-Noise for a Shaping Network to Reduce Side Lobes*

Frequently there are important advantages to be gained by trading a small amount of S/N for certain desirable signal properties. An important improvement in signal characteristics is obtained when the relative side-lobe strengths are reduced, and it will be found that substantial reduction in side lobes may be obtained at only a very small cost in S/N.

Both the active and passive cases are treated simultaneously by assuming the shaping-network input signal to be specified by

$$(\text{sinc } \Delta t)\ e^{2\pi i(f_0 t - kt^2/2)}.$$

To obtain results pertaining to the passive generation, it suffices to set $k = 0$, which is effectively equivalent to an infinite dispersion factor.

The specific shaping or weighting network to be studied is gaussian in shape and given by

$$\tilde{Y}_{\text{gau}}(f) = e^{-\pi G(f-f_0)^2}. \tag{51}$$

Here, G is a parameter related to the amount of taper introduced by $\tilde{Y}_{\text{gau}}(f)$. It is more convenient to use two other parameters, L and α, which are linearly related to G:

$$\begin{aligned} L &= 20 \log_{10} e^{\pi G \Delta^2/4} \\ &= \frac{20\pi}{4} (\log_{10} e)(G\Delta^2) \approx 6.83\, \alpha, \end{aligned} \tag{52a}$$

where

$$\alpha \equiv G\Delta^2. \tag{52b}$$

L represents the loss in decibels, imposed by $\tilde{Y}_{\text{gau}}(f)$, at the band edges, i.e. when $f = f_0 \pm \Delta/2$. Equation (51) may be rewritten in terms of $x = f/\Delta$ and $\alpha = G\Delta^2$:

$$\tilde{Y}_{\text{gau}}(x) = e^{-\pi\alpha(x-x_0)^2}. \tag{53}$$

For convenience, the convolution theorem $\epsilon_{\text{gau}}(y) = Y_{\text{gau}}(y) * \epsilon(y)$, will be expressed directly in natural units. For the input $\epsilon(y)$, a normalized response of the linear delay equalizer is employed:

$$\epsilon(y) = (\text{sinc } y) e^{2\pi i(x_0 y - y^2/2D)}. \tag{54}$$

The network impulse response, $Y_{\text{gau}}(y)$, becomes

$$\begin{aligned} Y_{\text{gau}}(y) &= \int_{-\infty}^{\infty} \tilde{Y}_{\text{gau}}(x) e^{2\pi i x y}\, dx \\ &= \frac{1}{\sqrt{\alpha}} e^{2\pi i x_0 y - \pi y^2/\alpha}. \end{aligned} \tag{55}$$

If this result is combined with (54), the modified complex response is given by

$$\epsilon_{\text{gau}}(y) = \frac{1}{\sqrt{\alpha}} e^{2\pi i x_0 y} \int_{-\infty}^{\infty} (\text{sinc } z) e^{-\pi[(y-z)^2/\alpha + iz^2/D]}\, dz. \tag{56a}$$

The integral in (56a) may be evaluated, and yields the following result for $\epsilon_{gau}(y)$:

$$\epsilon_{gau}(y) = \frac{1}{2\sqrt{\alpha}} e^{2\pi i x_0 y - \pi y^2/\alpha} \cdot \left\{ \operatorname{erf}\left[\sqrt{\frac{\pi}{c}} \left(\frac{1}{2} - \frac{iy}{\alpha} \right) \right] + \operatorname{erf}\left[\sqrt{\frac{\pi}{c}} \left(\frac{1}{2} + \frac{iy}{\alpha} \right) \right] \right\}, \tag{56b}$$

where c is a complex constant defined by

$$c = \frac{1}{\alpha} + \frac{i}{D}, \tag{57}$$

and "erf" denotes the usual error function. The peak signal, required for a S/N study, is attained at $y = 0$. Since

$$\int_{-\infty}^{\infty} \operatorname{sinc}^2 y \, dy = 1,$$

the input energy in (54) has the value 1. Therefore, the peak normalized signal power becomes

$$S_{gau} = | \epsilon_{gau}(0) |^2 = \frac{1}{\alpha} \left| \operatorname{erf}\left(\frac{1}{2} \sqrt{\frac{\pi}{c}} \right) \right|^2. \tag{58}$$

The noise power may be defined as

$$N_{gau} = \int | \hat{Y}_{gau}(x) |^2 \, dx, \tag{59}$$

when expressed in terms of the natural frequency variable. If (53) and (59) are combined, one obtains $N_{gau} = 1/\sqrt{2\alpha}$. From (57) and (58), then,

$$(S/N)_{gau} = \sqrt{\frac{2}{\alpha}} \left| \operatorname{erf}\left(\frac{1}{2} \sqrt{\frac{\alpha D \pi}{D + i\alpha}} \right) \right|^2. \tag{60}$$

In the case of ideal passive generation the *effective* value of D is infinity in (54) and, therefore, also in (60). Under these circumstances, (60) becomes

$$(S/N)_{gau} = \sqrt{\frac{2}{\alpha}} \left[\operatorname{erf}\left(\frac{1}{2} \sqrt{\pi\alpha} \right) \right]^2. \tag{61}$$

Actually this result is also quite accurate in the "active" case when D remains finite. In practice, a typical value of L, the loss in decibels introduced at the band edges, is of the order of 15–25 db; it follows from (52a) that $\alpha \approx 3$. For high-compression Chirp systems, whose D values are 100

and greater, the S/N in an active generation is well represented by (61). Fig. 9 shows a plot of the S/N degradation for a range of different gaussian weighting networks. It is noticed that there is a minimum loss of about 0.5 db when $L \approx 8.6$ db. A detailed study of the full response signal, (56b), has been made for the passive case; the results are approximately correct for actively generated pulses. Two gross signal features have been studied: (a) the output pulse width measured at the 3-db level and (b) the relative level between the main signal peak and the first adjacent side-lobe peak. The results of this study are summarized in Figs. 10(a) and 10(b). Qualitatively, as the loss introduced at the band edges by the gaussian network is increased, there is (a) a corresponding increase in pulse width and (b) an increase in side-lobe discrimination. The marked improvement in side-lobe discrimination is well worth the slight pulse-width increase and the small cost in S/N. This example illustrates that the maximum S/N attained only by a matched filter is, in reality, a very broad maximum. This general principle is exploited in Section 3.4.2, where the primary aim of weighting networks will be to improve signal properties using other networks more efficient than the simple gaussian one studied here.

The gaussian case does provide a qualitative picture of the effect produced by the residual FM in the active generation scheme. Suppose

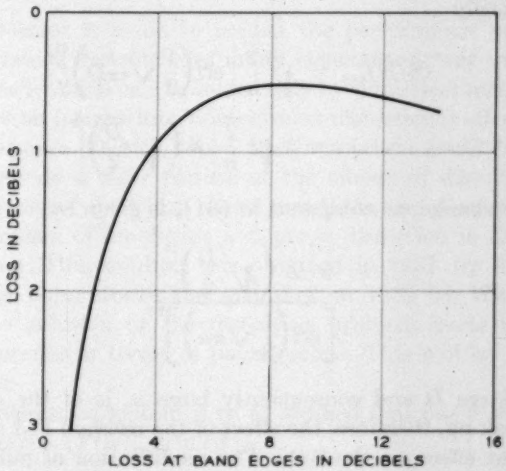


Fig. 9 — Degradation in S/N from ideal maximum when the rectangular Chirp signal passes through the delay equalizer and a smooth, gaussian-taper filter, which introduces a loss of L db at the band edges.

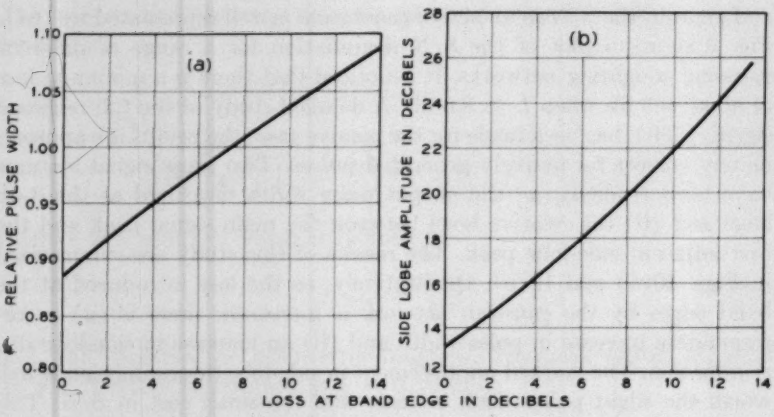


Fig. 10 — (a) Half-power pulse width increase vs. L, the gaussian filter band edge loss. (b) The relative amplitude between the maximum of the first adjacent side lobe and the central maximum of the output signal following shaping by a gaussian filter.

one was so intent on reducing the side-lobe level that L, and consequently α, was increased to a point where (61) no longer approximated (60). In fact, consider the other extreme where $\alpha \gg D$, so that (60) becomes essentially

$$(S/N)_{gau} = \sqrt{\frac{2}{\alpha}} \left| \operatorname{erf}\left(\frac{1}{2}\sqrt{i\pi D}\right) \right|^2 = 2\sqrt{\frac{2}{\alpha}} \left| Z\left(\sqrt{\frac{D}{2}}\right) \right|^2. \tag{62}$$

The relative change, as compared to (61), is given by

$$\frac{2\left| Z\left(\sqrt{\frac{D}{2}}\right) \right|^2}{\left[\operatorname{erf}\left(\frac{1}{2}\sqrt{\pi\alpha}\right)\right]^2},$$

which, for large D and consequently large α, is of the order of one. Crudely speaking, therefore, the effect of the residual FM seems to have no significant effect on the S/N. The modification of pulse width and side-lobe level by the residual FM would require a detailed consideration of the exact signal in (56b). This study is not warranted in view of the many advantages inherent in passive generation.

3.4 *A Study of Departures from Ideal Chirp Systems*

In this section, deviations from the ideal system behavior will be analyzed. These deviations will include the planned departures, such as the weighting network to reduce side lobes, and random deviations introduced by imperfect system components. A unified approach has been developed to predict the effects of both types of system irregularities.

The first topic, treated in Section 3.4.1, is the basic analytical tool, "paired-echo theory", which is used in all of the analysis on system distortions. The theory of spectrum weighting, Section 3.4.2, is then presented with the aid of paired-echo theory. The next subject in the sequence, Section 3.4.3, is a consideration of quadratic phase distortion or improper equalization of the FM transmitted pulse in the receiver. The last topic, Section 3.4.4, presents a treatment of the effects of moving targets on the collapsed-signal envelope characteristics. Although the effect of moving targets is not strictly a system distortion, the analytical investigation will also be presented with the aid of the paired-echo theory developed in Section 3.4.1. For this reason, it is natural to include the effects of moving targets in the present section.

3.4.1 *Paired-Echo Concept*

When an attempt is made to predict the performance of an actual Chirp radar system that includes many elements, a very unwieldy integral is obtained, which can be solved only by numerical methods. This, of course, gives no insight into how various distortional effects perturb the system response. It was hoped that a solution could be obtained that would provide a clear picture of the effects of distortion terms. The characteristics of the Chirp signal permit the use of a previous solution to the problem of amplitude and phase distortion in linear transmission systems. This solution was obtained in 1931 by MacColl[8] of Bell Telephone Laboratories and amplified in 1939 by Wheeler[13] and Burrows.[14] The solution of the distortion problem leads to a result which is interpreted in terms of paired echos. This will be clarified in the next section.

A linear transmission system is to be studied that has a steady-state transfer admittance, $[\tilde{Y}(\omega)]$, defined as

$$\tilde{Y}(\omega) = A(\omega)\epsilon^{jB(\omega)}, \tag{63}$$

where, as usual $\omega = 2\pi f$. The steady-state amplitude response is $A(\omega)$ and the steady-state phase characteristic is $B(\omega)$. In an ideal system,

$A(\omega)$ would be a constant independent of frequency and $B(\omega)$ would increase linearly with frequency. The physical system will exhibit variations from the ideal behavior that can be described by a Fourier series expansion about the frequency band of interest. This leads to a description of the system by the following equations:

$$A(\omega) = a_0 + \sum_n a_n \cos nc\omega, \tag{64}$$

$$B(\omega) = b_0\omega + \sum_n b_n \sin nc\omega. \tag{65}$$

If all a_n and b_n were zero except a_0 and b_0, the above equations would describe an ideal transmission system. Any distorting influence arising from a passive element can be described by these equations, as can active elements operated in their linear regions. Thus, all of the elements of a linear system can be represented in these terms.

MacColl's analysis[8] considers one term of the Fourier expansion given by (64) and (65). The steady-state amplitude and phase characteristic is given by:

$$A(\omega) = a_0 + a_1 \cos c\omega, \tag{66}$$

$$B(\omega) = b_0\omega + b_1 \sin c\omega. \tag{67}$$

Since a linear system is being considered, superposition will apply; therefore, the resultant output for an input $E(t)$ can be obtained as the sum of the responses to the various terms of the Fourier series expansion. The analysis obtains an output signal, $I(t)$, for an input signal, $E(t)$, to a system having the characteristics given in (66) and (67):

$$\begin{aligned} I(t) = {} & a_0 J_0(b_1) E(t + b_0) + J_1(b_1) \\ & \cdot \left[\left(a_0 + \frac{a_1}{b_1} \right) E(t + b_0 + c) - \left(a_0 - \frac{a_1}{b_1} \right) E(t + b_0 - c) \right] \\ & + J_2(b_1) \left[\left(a_0 + \frac{2a_1}{b_1} \right) E(t + b_0 + 2c) + \left(a_0 - \frac{2a_1}{b_1} \right) \right. \\ & \left. \cdot E(t + b_0 - 2c) \right] + J_3(b_1) \left[\left(a_0 + \frac{3a_1}{b_1} \right) E(t + b_0 + 3c) \right. \\ & \left. - \left(a_0 - \frac{3a_1}{b_1} \right) E(t + b_0 - 3c) \right] + \cdots . \end{aligned} \tag{68}$$

The functions $J_0(b_1)$, $J_1(b_1)$, $J_2(b_1)$, $\cdots$ are the usual Bessel functions; the first four are shown graphically in Fig. 11.

The solution given in (68) provides some insight into the perturba-

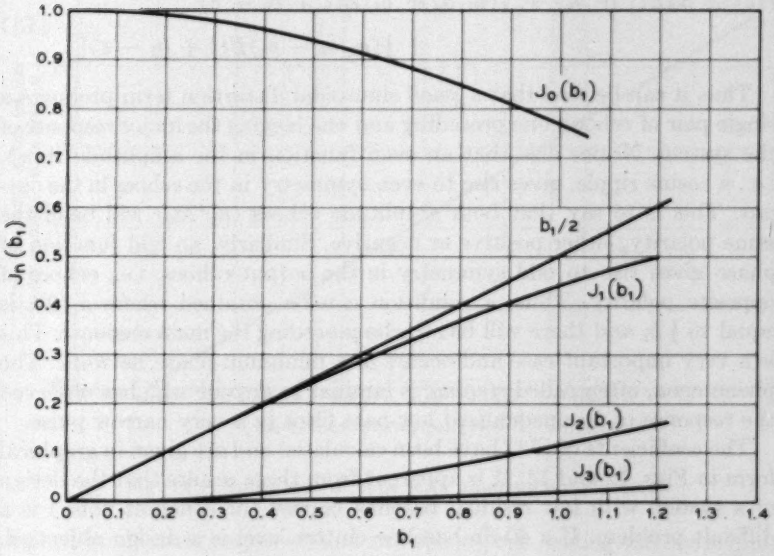


Fig. 11 — Bessel functions.

tions produced by distortion in a physical system. It can be seen that, with only one sinusoidal distortion term, the output is composed of a delayed replica of the input signal, $E(t + b_0)$, modified in amplitude by the coefficient, $a_0 J_0(b_1)$. The output also has an infinite series of terms — echoes that occur in pairs, one preceding and one lagging the major response — whose amplitude diminishes according to the Bessel function coefficients and whose separation from the major response is proportional to the order of the Bessel function. In a well-designed system, the coefficients a_1 and b_1 would be small. This leads to the following approximation for the Bessel functions:

$$J_0(b_1) = 1, \tag{69}$$

$$J_1(b_1) = \tfrac{1}{2}b_1, \tag{70}$$

$$J_n(b_1) = 0, \qquad \text{for } n > 1. \tag{71}$$

These approximations apply when

$$b_1 < 0.4 \text{ radians}. \tag{72}$$

This leads to a considerably simpler expression for the output, $I(t)$:

$$I(t) \cong a_0[E(t + b_0) + \tfrac{1}{2}(a_1/a_0 + b_1)E(t + b_0 + c) + \tfrac{1}{2}(a_1/a_0 - b_1)E(t + b_0 - c)]. \quad (73)$$

Thus it can be seen that a *small* sinusoidal distortion term produces a single pair of echoes, one preceding and one lagging the major response of the system. Notice also that an even function in the amplitude $A(\omega)$, i.e., a cosine ripple, gives rise to even symmetry in the echoes in the output. This is to say that both amplitude echoes ($a_1/2a_0$) will be of the same polarity, either positive or negative. Similarly, an odd function of phase gives rise to odd symmetry in the output echoes, i.e., echoes of opposite polarity. Thus, a condition can be obtained where $a_1/2a_0$ is equal to $\frac{1}{2}\ b_1$ and there will be no echo preceding the main response. This is a very important case and occurs in a minimum-phase network. The phenomena, often called *ringing*, is familiar to anyone who has observed the response of an unequalized low-pass filter to a very narrow pulse.

The coefficients of (73) have been calculated and are given in graphical form in Figs. 12 and 13. It is apparent from these results that the design of a system with low residual baseline clutter (or coherent noise) is a difficult problem. If a 40-db baseline clutter level is a design objective, each component, a_n or b_n, of the Fourier series expansion of (64) and (65) would have to be kept below the following low value for a minimum phase condition:

$$20 \log_{10}\left(1 + \frac{a_n}{a_0}\right) < 0.085 \text{ db}, \quad (74)$$

$$b_n < 0.57 \text{ degrees}. \quad (75)$$

Thus, it is seen that a 40-db clutter level is indeed extremely difficult to attain.

3.4.2 *Frequency Weighting to Reduce Side-Lobe Levels*

3.4.2.1 *Choice of Weighting Scheme.* Before proceeding into a detailed discussion of the weighting scheme employed in the present Chirp radar system, a simplified qualitative picture of the mechanics of weighting will be presented; this will permit a clearer understanding of the weighting process. The paired-echo theory of the preceding section will be used in the simplified description.

If a network with a raised-cosine amplitude response (as in Fig. 14) and a linear phase characteristic is excited with a sinc Δt input signal, the output signal, shown in Fig. 15, can be derived by paired-echo theory. The resultant output signal is the sum of a main delayed replica of the

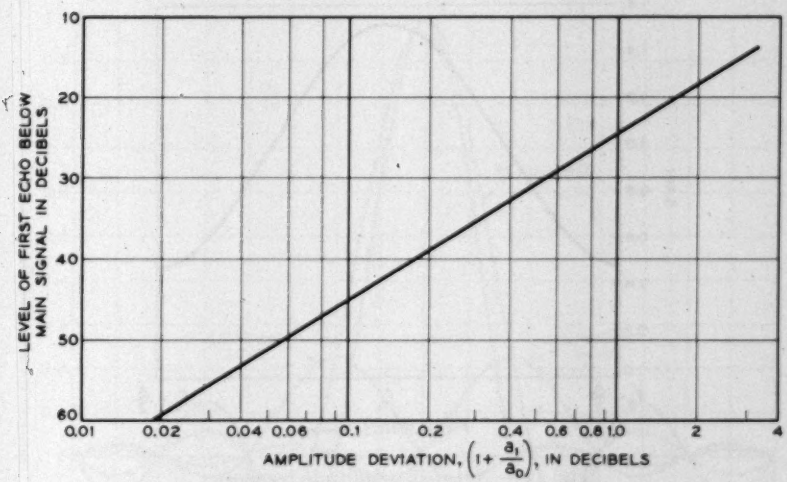


Fig. 12 — Amplitude of echo with amplitude distortion.

input signal (main response) and two echoes, one leading and one lagging the main response. It is seen that the amplitude of the echoes is equal to $a_1/2a_0$ and the time displacement of the echoes with respect to the main response is equal to the period of the cosine ripple, $1/\Delta$. The sum of the main response and the two echoes gives the resultant output signal, which

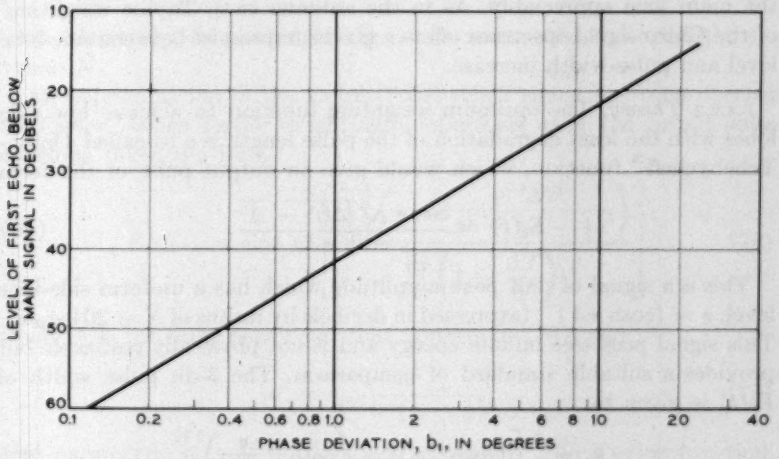


Fig. 13 — Amplitude of echo with phase distortion.

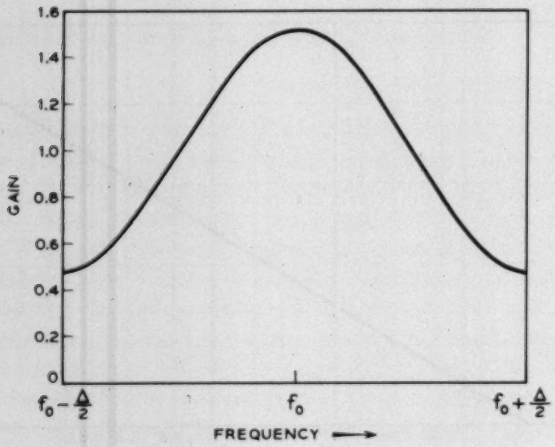


Fig. 14 — Network amplitude response.

has the desired lower side lobes at the expense of a slightly wider main pulse and a small loss in signal-to-noise ratio over the matched filter optimum.

The ideal weighting scheme should provide maximum attenuation of the sidelobes with minimum broadening of the main pulse. An analogous situation occurs in antenna theory with a line source. The antenna designer attempts t minimize the spatial side lobes without broadening the main lobe appreciably. As in the antenna case, Taylor weighting[9] of the Chirp signal spectrum offers a good compromise between side-lobe level and pulse-width increase.

3.4.2.2 *Theory.* The optimum weighting function to achieve low side lobes with the least degradation of the pulse length is a so-called Dolph-Tchebycheff[15] function, which would give an output pulse of the form

$$E_0(t) = \frac{\cos \pi \sqrt{(\Delta t)^2 - A^2}}{\cosh \pi A}. \tag{76}$$

This is a signal of unit peak amplitude which has a uniform side-lobe level, $\eta = (\cosh \pi A)^{-1}$ (expressed in decibels by means of $\bar{N} = 20 \log_{10} \eta$). This signal possesses infinite energy and is not physically realizable but provides a suitable standard of comparison. The 3-db pulse width of $E_0(t)$ is given by

$$\beta_0 = \frac{2}{\pi}\left[(\cosh^{-1}\eta)^2 - \left(\cosh^{-1}\frac{\eta}{\sqrt{2}}\right)^2\right]^{\frac{1}{2}}. \tag{77}$$

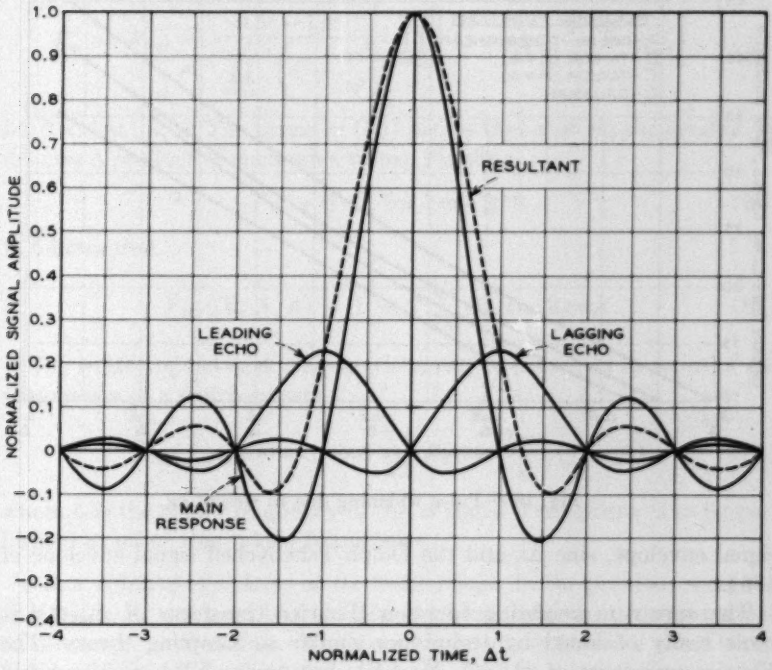


Fig. 15 — Weighting mechanics.

This relationship is plotted in Fig. 16 as normalized pulse length, β_0, versus $\bar{N}$.

An approximation of the Dolph-Tchebycheff signal has been derived by Taylor.[9] When Taylor weighting functions are used, the output pulse envelope takes the form

$$E_{tay}(t) = \text{sinc } \Delta t \left[\frac{\prod_{n=1}^{\bar{n}-1} \left(1 - \frac{\sigma^{-2}\Delta^2 t^2}{A^2 + (n - \frac{1}{2})^2} \right)}{\prod_{n=1}^{\bar{n}-1} \left(1 - \frac{\Delta^2 t^2}{n^2} \right)} \right], \tag{78}$$

where A is the same parameter that appears in (76), and

$$\sigma^2 \equiv \frac{\bar{n}^2}{A^2 + (\bar{n} - \frac{1}{2})^2}. \tag{79}$$

The parameter, $\bar{n}$, will be described presently. The Taylor function, $E_{tay}(t)$, of (78) is a compromise between the unweighted Chirp output

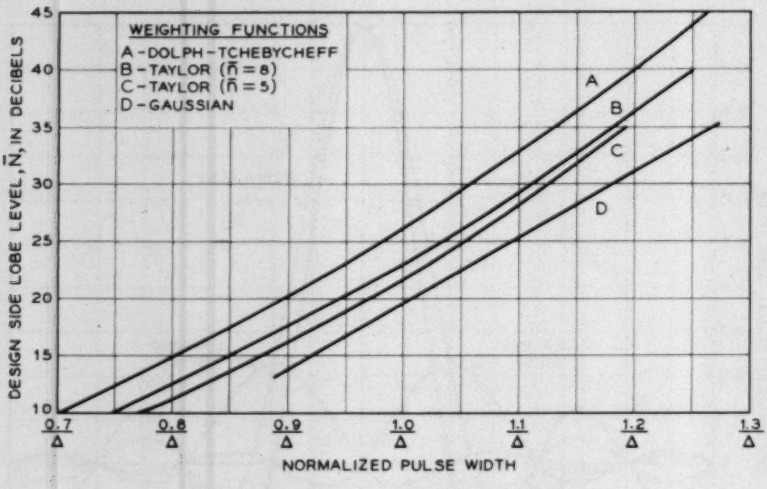


Fig. 16 — Pulse widening due to weighting.

signal envelope, sinc Δt, and the Dolph-Tchebycheff signal envelope of (76).

The spectrum-weighting function [Fourier transform of $E_{\text{tay}}(t)$] is most easily obtained by techniques similar to sampling theory. The Fourier transform of $E_{\text{tay}}(t)$, $\tilde{Y}_{\text{tay}}(X)$, is zero for $|X| > 0.5$, where $X = (f - f_0)/\Delta$:

$$E_{\text{tay}}(t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{Y}_{\text{tay}}(X, A, \bar{n}) e^{2\pi i X \Delta t} dX. \tag{80}$$

Sampling theory suggests the following substitution:

$$\tilde{Y}_{\text{tay}}(X, A, \bar{n}) = \sum_{-\infty}^{\infty} F_m e^{-2\pi i m X}. \tag{81}$$

From this, one finds

$$E_{\text{tay}}(t) = \sum_{-\infty}^{\infty} F_m[\text{sinc}\ (\Delta t - m)]. \tag{82}$$

It also follows that

$$F_m = F(m, A, \bar{n}). \tag{83}$$

The coefficients, F_m, can be obtained as follows:

$$F_0 = 1, \tag{84}$$

$$F_m = F_{-m} = \frac{0.5(-1)^{m+1}}{\prod_{n=1}^{(\bar{n}-1)'} \left(1 - \frac{m^2}{n^2}\right)} \prod_{n=1}^{\bar{n}-1} \left[1 - \frac{\sigma^{-2}m^2}{A^2 + (n - \frac{1}{2})^2}\right], \quad (85)$$

for $0 < |m| < \bar{n}$. The prime in (85) means that $n \neq m$; the symbol $\prod$ denotes a product of successive terms. Finally,

$$F_m = 0 \qquad \text{for } |m| \geqq \bar{n}. \quad (86)$$

It follows that

$$\bar{Y}_{\text{tay}}(X, A, \bar{n}) = 1 + 2 \sum_{m=1}^{\bar{n}-1} F_m \cos 2\pi m X. \quad (87)$$

The weighting network characteristic can be expressed in terms of the parameters of a Chirp system:

$$\bar{Y}_{\text{tay}}(f, A, \bar{n}) = 1 + 2 \sum_{m=1}^{\bar{n}-1} F_m \cos \frac{2\pi m}{\Delta} (f - f_0), \quad (88)$$

where f_0 is the center frequency of the IF signal if weighting is to be performed at other than video frequencies.

3.4.2.3 *Effect of Weighting on the Pulse Shape.* In the previous section a parameter, $\bar{n}$, was used but has not yet been defined; it is appropriate to discuss it at this time. When a comparison is made of the Dolph-Tchebycheff and Taylor pulse envelopes for the same design side-lobe level (see Fig. 17), there are several important differences: The Taylor pulse width, expressed at the 3-db pulse width, is

$$\beta = \sigma\beta_0, \quad (89)$$

where β_0 is the Dolph-Tchebycheff 3-db pulse width given in (77) and the quantity σ is defined in (79). An important analytical difference is the location of the zeros of the two functions. The Dolph-Tchebycheff envelope has all of its zeros occurring at nonintegral values of the argument Δt, while the Taylor envelope has two distinct regions: a region of nearly uniform side lobes where the zeros occur at nonintegral values of the argument $|\Delta t| \leqq \bar{n} - 1$, and a region of decaying side lobes where the zeros always occur at integers $|\Delta t| \geqq \bar{n}$. The central region of near-uniform side lobes approximates the Dolph-Tchebycheff behavior. It is essential to have the remote side lobes decay to keep their energy content down.

From the previous discussion it is easy to see the significance of the parameter $\bar{n}$, since it gives a bound on the region of approximation to Dolph-Tchebycheff behavior. Higher values of $\bar{n}$ mean a closer approxi-

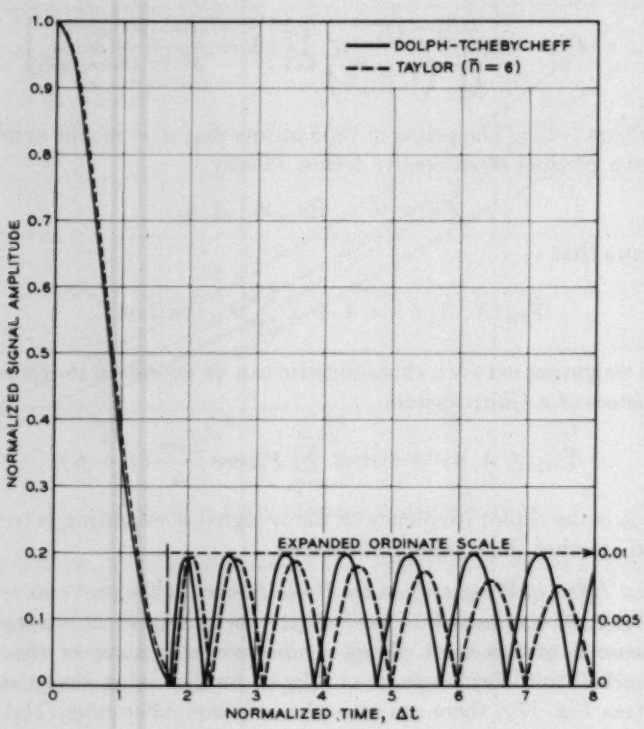


Fig. 17 — Comparison of Taylor and Dolph-Tchebycheff signal envelopes for $\tilde{N} = 40$ db.

mation. The upper bound in the antenna case is determined by supergaining but, for the Chirp case, it is not desirable to use any larger value of $\bar{n}$ than the minimum permissible. When the minimum value of $\bar{n}$ is chosen, the required weighting characteristic is easily obtained with a bandpass filter. This is evident from the results presented in Figs. 18 and 19, which are plots of the required loss functions for design side-lobe levels of 40 and 47.5 db for various values of $\bar{n}$, ranging upward from the minimum permissible value in both cases. In order to obtain the minimum permissible value, Taylor employs the rule that, as $\bar{n}$ is increased by one, the value of σ should not increase.

The required loss functions for side-lobe levels of 34, 37.5, 40 and 47.5 db have been calculated for the minimum value of $\bar{n}$ in each case and are plotted in Fig. 20. The resultant pulse envelopes are shown in Figs. 21 through 24. The side-lobe levels have been amplified in some of the figures to show the details.

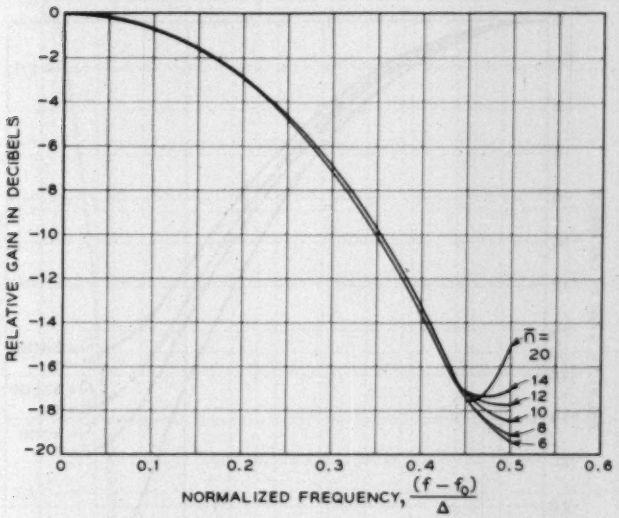


Fig. 18 — Taylor weighting characteristics for $\bar{N}$ = 40 db.

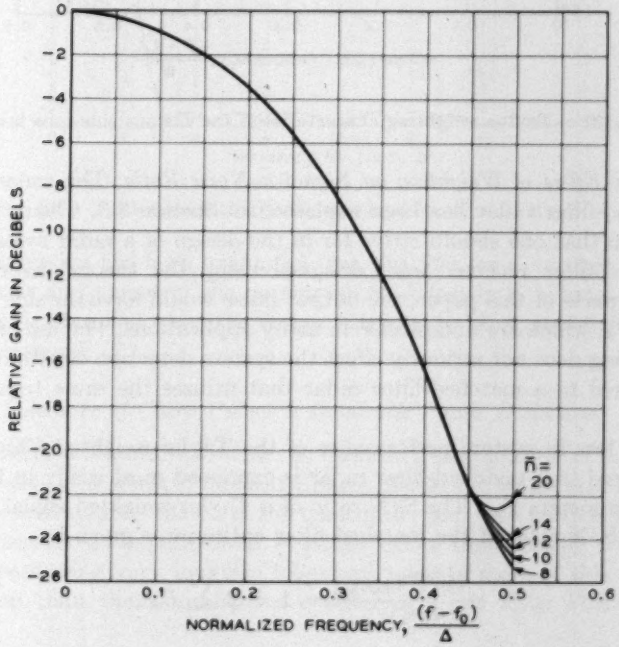


Fig. 19 — Taylor weighting characteristics for $\bar{N}$ = 47.5 db.

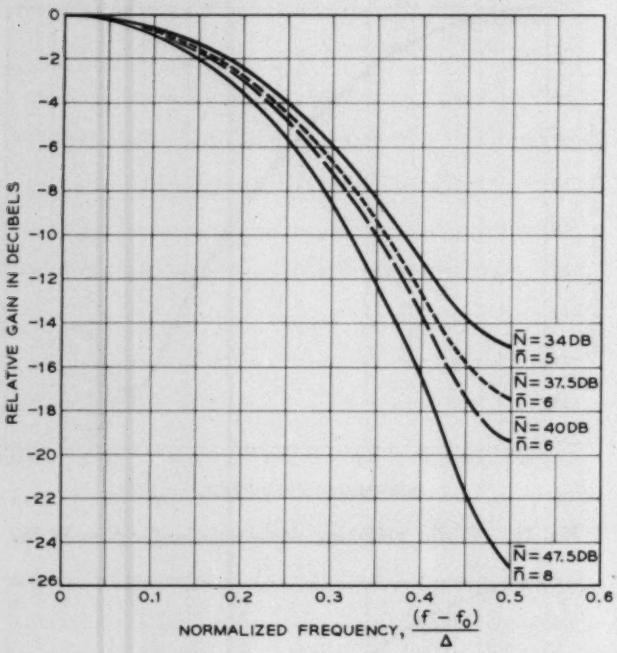


Fig. 20 — Taylor weighting characteristics for various side-lobe levels.

3.4.2.4 *Effect of Weighting on Signal-to-Noise Ratio.* The concept of a matched-filter radar has been explained in Section 3.3. Clearly, this is the goal that one should strive for in the design of a radar system. If a matched-filter receiver were designed using the theory described in earlier parts of this paper, the output pulse would have the side lobes of $(\sin z)/z$, which are undesirable in many applications. The use of Taylor weighting does not seriously effect the system detection capability when compared to a matched-filter radar that utilizes the same transmitted energy.

The loss in system performance of the Taylor-weighted Chirp radar compared to a matched-filter radar is expressed most easily in terms of the parameters F_m. The S/N ratio of a Taylor-weighted signal relative to the S/N ratio of the matched filter optimum is given by

$$\frac{(S/N)_{\text{matched}}}{(S/N)_{\text{taylor}}} = 1 + 2\sum_{m=1}^{\bar{n}-1} F_m^2. \tag{90}$$

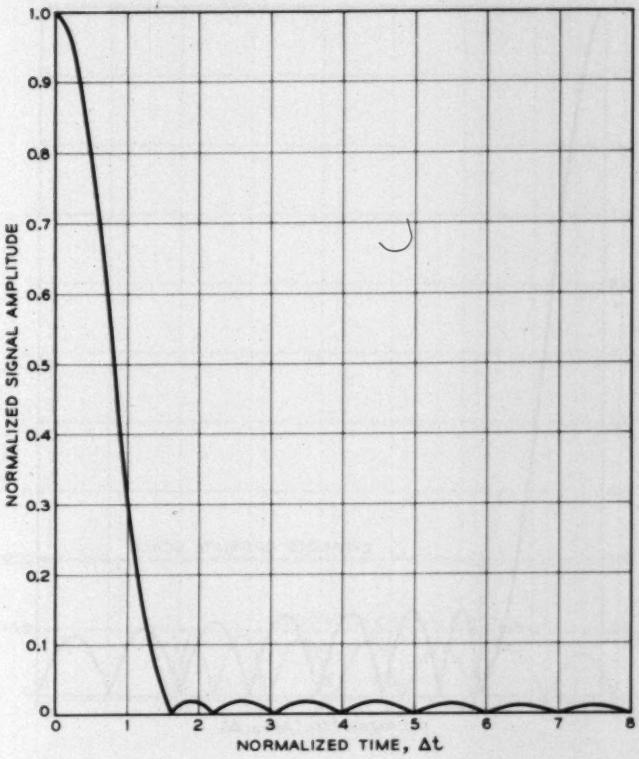


Fig. 21 — Taylor weighting pulse envelope for $\bar{N} = 34$ db, $\bar{n} = 5$.

This expression has been evaluated (see Fig. 25) for a number of side-lobe levels and values of the parameter $\bar{N}$. A loss in S/N ratio of only 1.455 db is obtained when the design side-lobe level, $\bar{n}$, is 47.5 db below the peak signal.

3.4.3 *Tolerance of Chirp Pulses to Quadratic Phase Distortion*

This section is concerned with the unwanted degradation of collapsed Chirp pulses as a result of quadratic phase distortion, which could be the result of an imperfect match of the transmitter and receiver delay-equalizer slopes. Some experimental evidence exists that heavily weighted pulses possessing very low side lobes can tolerate more of this type of distortion than their unweighted counterparts $(\sin z)/z$. This section

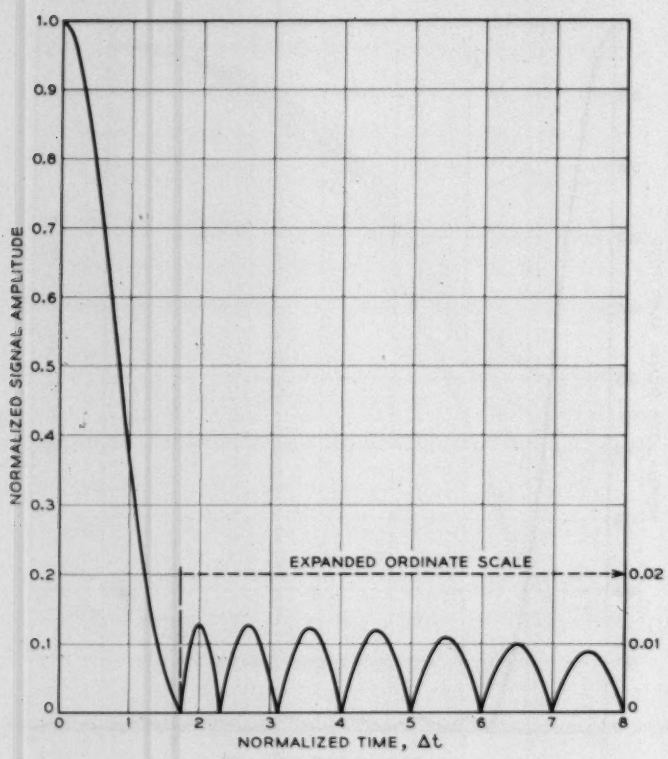


Fig. 22 — Taylor weighting pulse envelope for $\tilde{N} = 37.5$ db, $\bar{n} = 6$.

shows analytically that unweighted pulses are degraded more than heavily weighted pulses. The particular weighting function chosen for this analysis is slightly less efficient than those previously described (Section 3.4.2) but represents the Taylor-weighted optimum to a reasonable approximation. It is believed that the general results of this section are appropriate for any *heavily* weighted pulse.

3.4.3.1 *Analysis of Quadratic Phase Distortion.* The weighting scheme described in Section 3.4.2 used a network having a transfer admittance given by

$$\tilde{Y}_{tay}(f, A, \bar{n}) = 1 + 2 \sum_{m=1}^{\bar{n}-1} F_m \cos \frac{2\pi m}{\Delta} (f - f_0). \qquad (91)$$

This network, when excited by the collapsed signal,

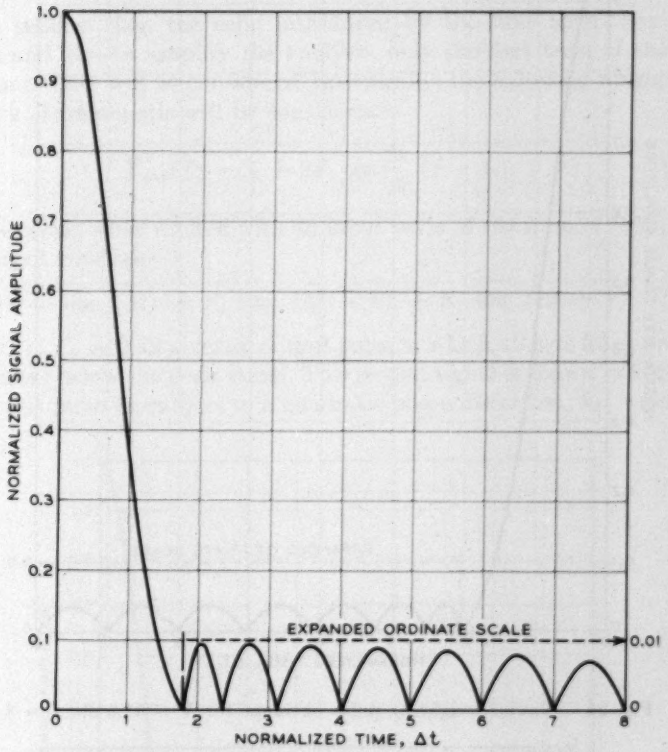


Fig. 23 — Taylor weighting pulse envelope for $\bar{N} = 40$ db, $\bar{n} = 6$.

$$e(t) = \cos 2\pi f_0 t \operatorname{sinc} \Delta t, \tag{92}$$

gives rise to an output signal envelope, $E_{tay}(t)$, where

$$E_{tay}(t) = \sum_{m=-(\bar{n}-1)}^{(\bar{n}-1)} F_m \operatorname{sinc} (\Delta t - m). \tag{93}$$

The desired weighting-network characteristic is described by (91), but the coefficients F_m decrease rapidly as m increases. In the particular case considered in this section, $\bar{N} = 40$ db and $\bar{n} = 6$ (the minimum permissible value), the following ratios exist:

$$\left|\frac{F_5}{F_1}\right| < \left|\frac{F_4}{F_1}\right| < \left|\frac{F_3}{F_1}\right| < \left|\frac{F_2}{F_1}\right| < 0.025. \tag{94}$$

Thus, it can be seen that the echoes introduced in $E_{tay}(t)$ by the second through fifth terms in the expansion in (91) are at least an order of mag-

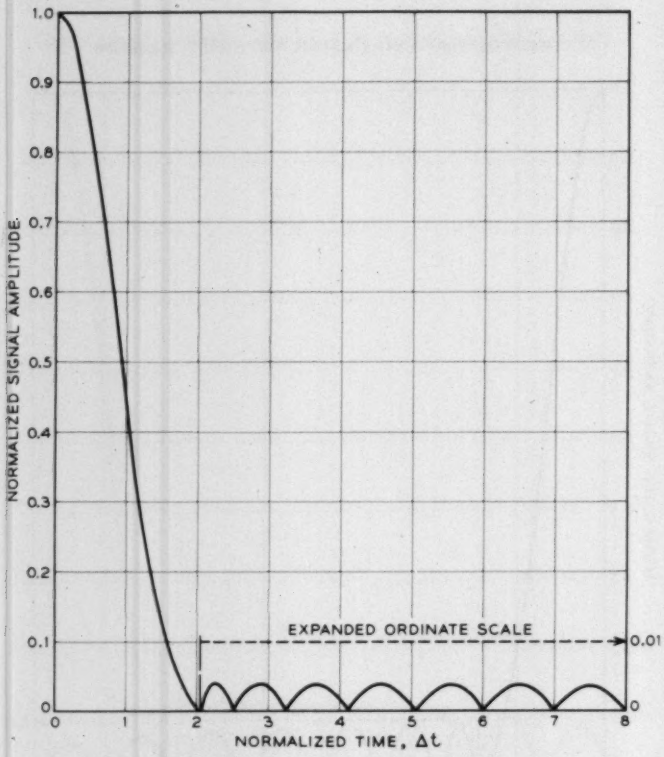


Fig. 24 — Taylor weighting pulse envelope for $\bar{N} = 47.5$ db, $\bar{n} = 8$.

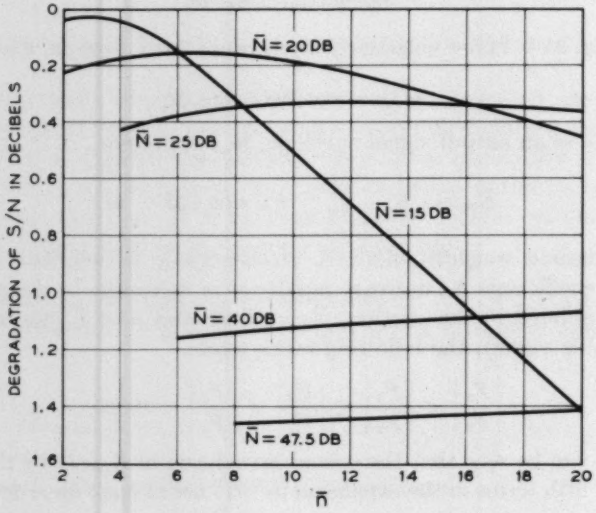


Fig. 25 — Signal-to-noise effects.

nitude smaller than the echo introduced by the first term. For this reason and also to simplify the analysis, only the first term of the expansion of (91) will be considered. Specifically, the following weighting-network characteristic will be considered:

$$\tilde{Y}'_{\text{tay}}(f) = 1 + 2F_1 \cos \frac{2\pi}{\Delta} (f - f_0). \tag{95}$$

This network, when excited with an input pulse of the form of (92), has the output envelope

$$E'_{\text{tay}}(t) = \text{sinc } (\Delta t) + F_1 \text{ sinc } (\Delta t - 1) + F_1 \text{ sinc } (\Delta t + 1). \tag{96}$$

A value of $F_1 = 0.42$ gives an output pulse in which all side lobes are 40 db or more below the peak signal. This output signal is shown in Fig. 26.

Let this pulse be subject to a quadratic phase distortion, Φ_d, where

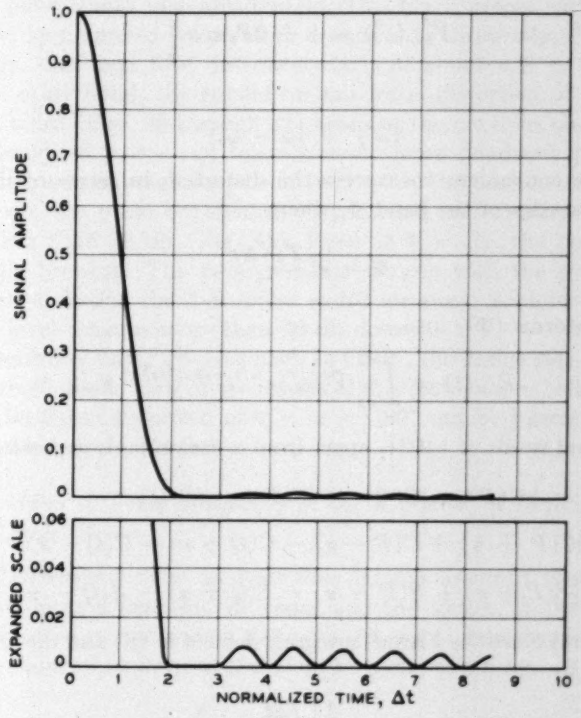


Fig. 26 — Quadratic phase distortion effects, weighted pulse, $\varphi = 0$.

$$\Phi_d \equiv -\tfrac{1}{2}T_0^2(\omega - \omega_0)^2. \tag{97}$$

This can be expressed as a linear delay distortion

$$t_d = -\frac{d\Phi}{d\omega} = T_0^2(\omega - \omega_0). \tag{98}$$

It should be emphasized that the linear delay in (98) represents a *distortion* over and above the *planned* linear delay introduced by the Chirp receiving network in (13).

The distorted output-pulse envelope is found by use of the inverse Fourier transform. Omitting constant values,

$$E_{\text{dis}}(t) = \int_{-\pi\Delta}^{\pi\Delta} \bar{Y}'_{\text{tay}}(\omega')e^{i\omega' t}e^{-i[T_0^2(\omega')^2]/2}\,d\omega', \tag{99}$$

where

$$\bar{Y}'_{\text{tay}}(\omega') = 1 + 2F_1 \cos\frac{\omega'}{\Delta}, \tag{100}$$

and

$$(\omega')^2 = (\omega - \omega_0)^2. \tag{101}$$

It is more convenient to express the distortion in terms of the phase shift at the edge of the band, Φ, where

$$\Phi = \pi^2\frac{T_0^2(\Delta)^2}{2}. \tag{102}$$

This transforms (99) into

$$E_{\text{dis}}(t) = \int_{-\pi\Delta}^{\pi\Delta} \bar{Y}'(\omega')e^{i\omega' t}e^{-i\Phi\omega'^2/\pi^2\Delta^2}\,d\omega'. \tag{103}$$

The general result of (103), apart from a constant, is expressed as

$$\begin{aligned}|E_{\text{dis}}(t)| = |\,&C(P) - C(Q) + iS(P) - iS(Q)\\ &+ F_1[C(P+\pi) + C(P-\pi) - C(Q+\pi) - C(Q-\pi)]\\ &+ jF_1[S(P+\pi) + S(P-\pi) - S(Q+\pi) - S(Q-\pi)]\,|,\end{aligned} \tag{104}$$

where C and S are the Fresnel integrals defined in (8) and the arguments are given by

$$P = \frac{1}{\sqrt{\pi}}\left(\frac{\Delta t}{2n} + n\right), \tag{105}$$

$$Q = \frac{1}{\sqrt{\pi}}\left(\frac{\Delta t}{2n} - n\right), \tag{106}$$

$$n = \sqrt{2\Phi}. \tag{107}$$

Notice that the echo structure of the signal is unchanged by the distortion process. The unweighted signal response is found very simply by setting F_1 equal to zero.

3.4.3.2 *Discussion of Results of Quadratic Phase Distortion Analysis.* The effect of quadratic phase distortion on weighted and unweighted pulses has been calculated according to (104), and the results are shown in Figs. 26 through 36. With no weighting, the side-lobe structure is appreciably distorted with a quadratic phase distortion as small as $\Phi = \pi/8 = 45°$ (see Fig. 34). Moreover, when $\Phi = 2\pi$ (Fig. 36), the peak signal is attentuated by more than 7 db, the pulse breaks up into two peaks, and the 3-db pulse length increases tenfold. This result agrees with the rule of thumb, long applied by antenna designers to uniformly illuminated aperatures, that near field distortion starts at about $\pi/8 < \Phi < \pi/4$.

On the other hand, the results for quadratic distortion of a heavily weighted pulse (Figs. 26 through 32) bear out the previous observation: heavily weighted pulses can tolerate much more quadratic phase distortion than can unweighted pulses. At $\Phi = \pi$, the peak attenuation is only 1.5 db, the pulse broadening 40 per cent, and the side-lobe level remains less than 38 db (Fig. 31). Even at $\Phi = 2\pi$, the pulse shape exhibits no breakup. The pulse remains smooth with the peak signal attenuated 4 db and the 3-db pulse width increased 2.3-fold; the peak side-lobe level remains more than 36 db down.

In present-day radar systems such as Chirp, the design side-lobe level is less than 40 db. In view of the results of this section, it is believed that a quadratic phase distortion of $\Phi = \pi = 180°$ can be tolerated at the edges of the frequency band.

3.4.4 *Problem of Fixed Weighting in the Reception of Doppler-Shifted Signals*

3.4.4.1 *Introduction.* The analysis thus far has assumed that there is no relative motion between the radar and the target. If the target is moving relative to the radar, the returned signal experiences a slight shift in frequency (known as the Doppler shift) given by

$$\nu = \frac{2v}{c} f_{RF}, \tag{108}$$

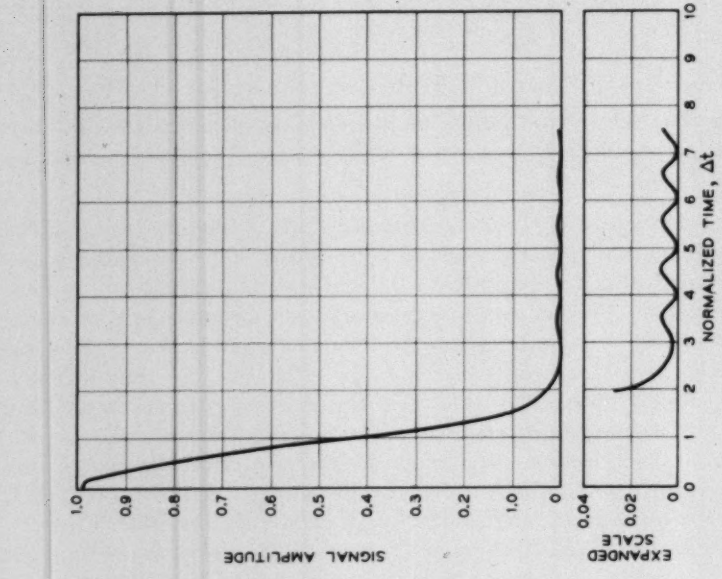


Fig. 27 — Quadratic phase distortion effects, weighted pulse, $\varphi = \pi/8$.

Fig. 28 — Quadratic phase distortion effects, weighted pulse, $\varphi = \pi/6$.

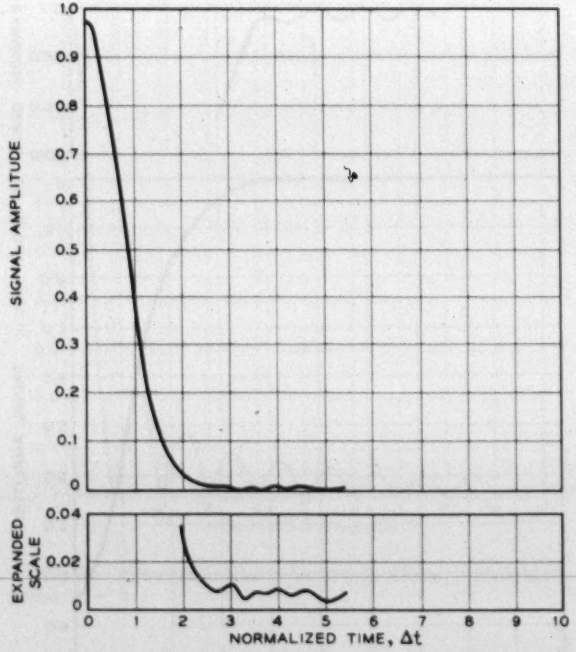


Fig. 29 — Quadratic phase distortion effects, weighted pulse, $\varphi = \pi/4$.

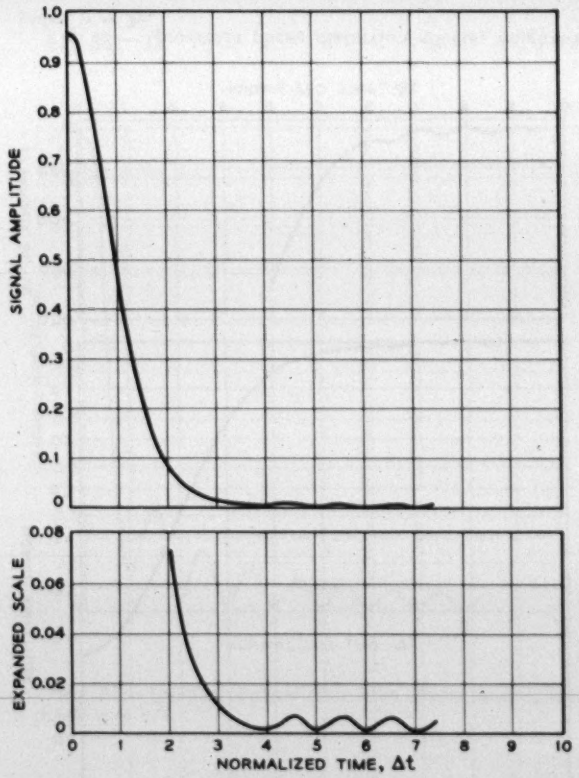


Fig. 30 — Quadratic phase distortion effects, weighted pulse, $\varphi = \pi/2$.

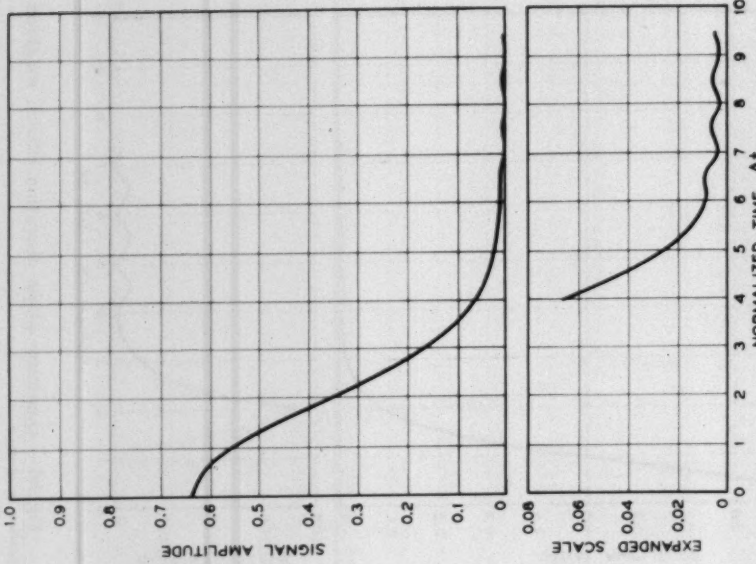


Fig. 32 — Quadratic phase distortion effects, weighted pulse, $\varphi = 2\pi$.

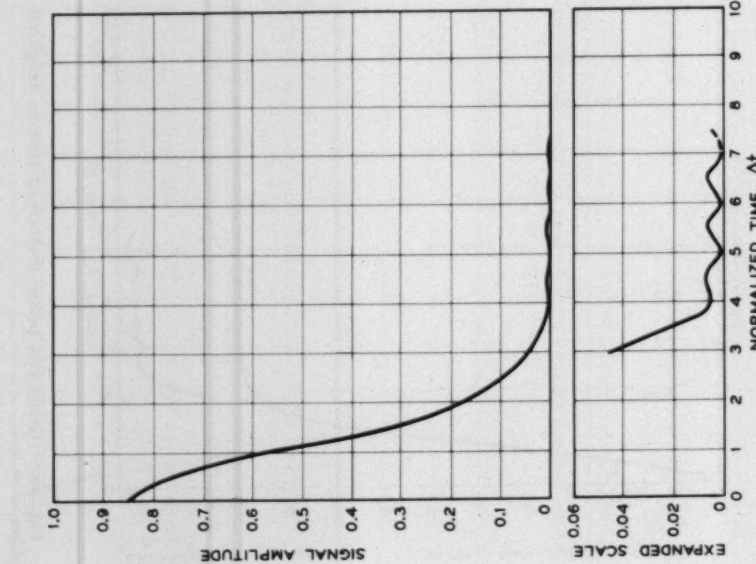


Fig. 31 — Quadratic phase distortion effects, weighted pulse, $\varphi = \pi$.

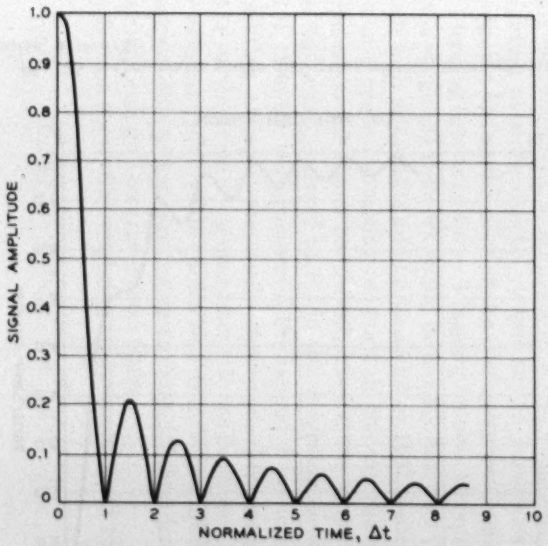


Fig. 33 — Quadratic phase distortion effects, unweighted pulse, $\varphi = 0$.

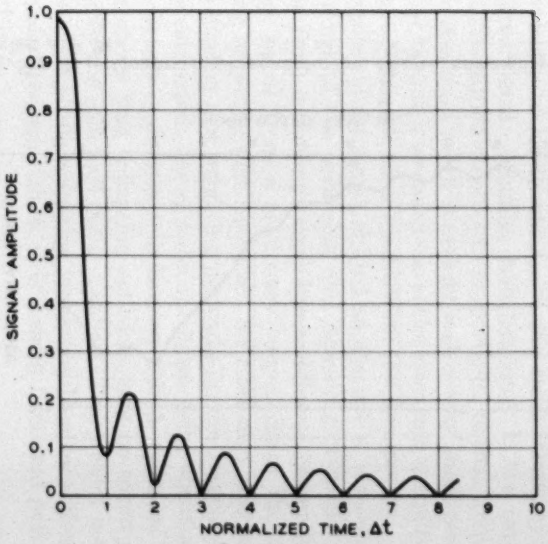


Fig. 34 — Quadratic phase distortion effects, unweighted pulse, $\varphi = \pi/8$.

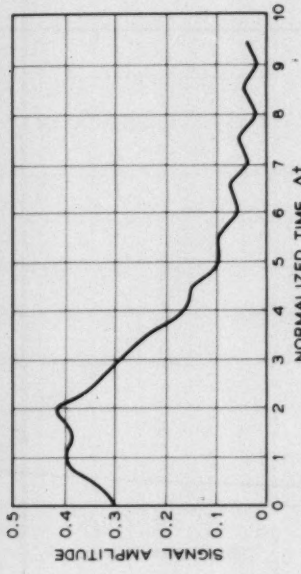


Fig. 36 — Quadratic phase distortion effects, unweighted pulse, $\varphi = 2\pi$.

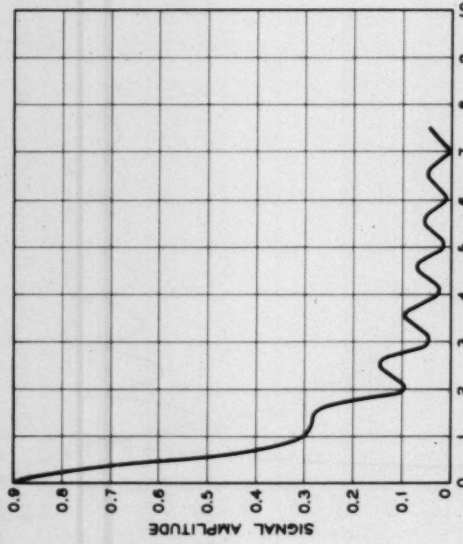


Fig. 35 — Quadratic phase distortion effects, unweighted pulse, $\varphi = \pi/2$.

where v is the relative radial velocity between the target and the radar, c is the velocity of light, and f_{RF} is the radio-frequency carrier of the transmitted signal. Actually, each frequency in the transmitted spectrum will be shifted an amount proportional to that frequency; in practice, the spectrum width is usually such a very small percentage of the RF carrier that it is assumed the entire spectrum is bodily translated an amount ν. Negligible error results from this assumption.

The radar return signal, shifted an amount ν, is modulated down to the IF range of the collapsing and weighting networks. The IF signal resulting from a moving target has a different center frequency than does the signal from a stationary target. Since the collapsing and weighting networks were optimized for stationary targets, the effect of these collapsing and weighting networks on the Doppler-shifted signals must be calculated. The effect of the linear delay equalizer on Doppler-shifted signals is discussed in Section 3.5. A principal result of that analysis shows that collapsing networks may be constructed that will not modify the collapsed pulse envelope. The present section considers only the effect of the weighting process on Doppler-shifted signals when the Doppler frequency shift, ν, is a small fractional part of the system bandwidth, Δ.

3.4.4.2 *Analysis.* The paired-echo explanation of the weighting process (Section 3.4.2) suggests another method of realizing the necessary weighting characteristic, in addition to a loss equalizer. The output of the weighting network can be viewed as the superposition of a large major response [delayed replica of the $(\sin z)/z$ input] pulse and a series of $\bar{n} - 1$ leading echoes and $\bar{n} - 1$ lagging echoes. An equivalent realization of the weighting network would be a delay line with $[2(\bar{n} - 1) + 1]$ taps, each with adjustable amplitude and 0 or 180° phase shift, brought together in a central summing network. If the tap spacing is equal to the reciprocal of the signal bandwidth, Δ, and the amplitude setting of the mth tap is proportional to F_m in the weighting network expansion [see (91)], then the performance of the tapped delay line is identical to the weighting network previously described. The tapped delay line method of synthesizing a loss characteristic has been used in the past,[16,17] and the resultant network is known as a transversal filter. The analysis and results of this section are not essentially different from the work of Bellows and Graham[18] on transversal filters.

As in Section 3.4.3, the weighting characteristic will be considered as

$$\bar{Y}'_{tay}(f) = 1 + 2F_1 \cos \frac{2\pi}{\Delta} (f - f_0). \tag{109}$$

For a moving target the mean echo frequency is

$$f_m = f_0 + \nu, \tag{110}$$

where ν has been defined in (108). The spectrum of the received signal is translated and will be constant in the range

$$f_m - \Delta/2 < f < f_m + \Delta/2, \tag{111}$$

and equal to zero elsewhere. With this fact, the output of the weighting network is proportional to

$$E'_{\text{tay}}(t) = \int_{f_m-\Delta/2}^{f_m+\Delta/2} \left[1 + 2F_1 \cos \frac{2\pi}{\Delta}(f - f_0)\right] e^{i2\pi ft}\, df. \tag{112}$$

An equivalent result is obtained by the passage of a signal with no Doppler shift through a weighting network with a slightly shifted center frequency. With this interpretation, (112) becomes

$$E'_{\text{tay}}(t) = \int_{f_0-\Delta/2}^{f_0+\Delta/2} \left[1 + 2F_1 \cos \frac{2\pi}{\Delta}(f - f_0 - \nu)\right] e^{i2\pi ft}\, df. \tag{113}$$

The envelope of this function is given by

$$\begin{aligned} |E'_{\text{tay}}(t)| = |&\operatorname{sinc} \Delta t + F_{11} \operatorname{sinc} (\Delta t + 1) + F_{11} \operatorname{sinc} (\Delta t - 1) \\ &+ iF_{12}[-\operatorname{sinc} (\Delta t + 1) + \operatorname{sinc} (\Delta t - 1)]|, \end{aligned} \tag{114}$$

where

$$F_{11} = F_1 \cos (2\pi\nu/\Delta), \tag{115}$$

and

$$F_{12} = F_1 \sin (2\pi\nu/\Delta). \tag{116}$$

For small values of ν/Δ,

$$\begin{aligned} F_{11} &\cong F_1, \\ F_{12} &\cong 2\pi\nu/\Delta. \end{aligned} \tag{117}$$

Therefore, the real component of the side lobes remains unchanged. The quadrature echoes become increasingly important as the percentage Doppler shift, ν/Δ, increases.

From the results of (114), it can be shown that Doppler shifts of less than 5 per cent of the system bandwidth, Δ, will cause very little degradation of the weighted output pulse. The breakover point for serious effects occurs for values of ν/Δ somewhat greater than 10 per cent. This result has been verified by an independent calculation of the output

pulse shape from a 47.5-db weighting network for a 10 per cent Doppler shift. This calculation (Fig. 37) shows a 0.2-db degradation of the peak pulse amplitude and a 13-db increase in the side-lobe level to 34.5 db.

3.5 *The Influence of Moving Targets on Chirp Systems*

This section continues the study of the effect of moving targets initiated in Section 3.4.4. Here attention is concentrated on the influence of moving targets on the process of *compression* as opposed to the study in 3.4.4 which was confined to the influence of moving targets on the process of *weighting*. In the study in this section of the effect of moving targets on a Chirp system it is initially assumed that only the linear-delay phase equalizer is present. With this simplified receiver characteristic, the quantitative effect of the Doppler shift may be readily found. Inspection of the receiver network characteristics in Fig. 7(b) reveals that the distortionless frequency translation caused by the Doppler shift results in a corresponding time translation of the network output

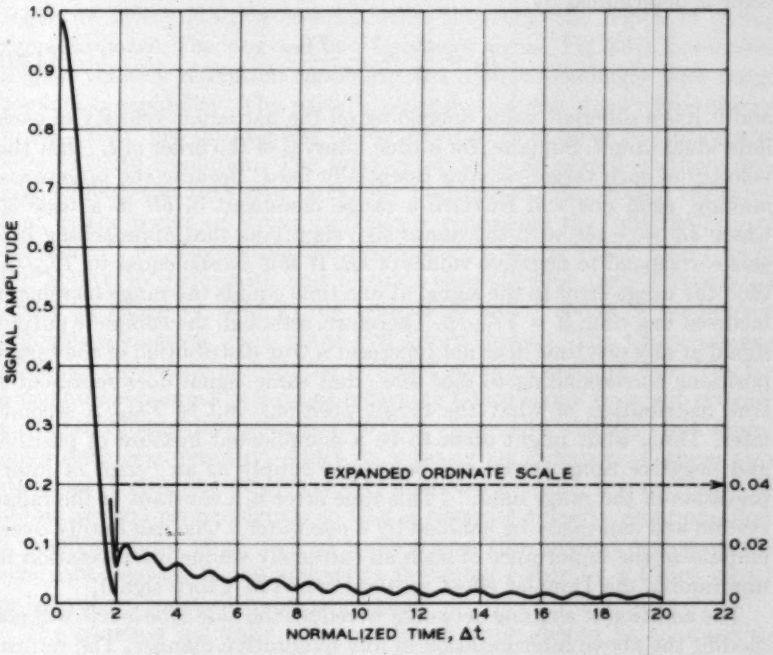


Fig. 37 — Weighting effects on Doppler-shifted signals.

signal. According to the particular network slope illustrated in Fig. 7(b), the response of an approaching target, with positive Doppler shift, will emerge *before* the response of a hypothetical stationary target at the same range. Likewise, the response of a receding target will lag the hypothetical stationary target response. The amount of time lead or lag may be determined from the following relation:

$$\frac{t_l}{\nu} = -\frac{T}{\Delta}, \tag{118}$$

where $t_l > 0$ denotes a time lag and $t_l < 0$ denotes a time lead. If (108) for ν is combined with (118), it follows that

$$t_l = -\frac{2vTf_{RF}}{c\Delta}. \tag{119}$$

It is important to point out that t_l is simply proportional to the relative velocity of the moving target. The time shift given by (119) will give rise to an error in the indicated range for every moving target. This range error is determined by

$$R_e = \frac{ct_l}{2} = -\frac{vTf_{RF}}{\Delta}, \tag{120}$$

and it has a different value depending on the particular velocity of each individual target. Suppose, for a time interval of the order of t_l, that the velocity of each target remains essentially fixed. Because the targets are moving, each one will traverse a range increment of δR in a time δt, where $\delta R = -v\delta t$, with the minus sign signifying that approaching targets correspond to negative values of δR. If δt is chosen equal to Tf_{RF}/Δ, then the range error in the signal at one time equals the range increment incurred in a time $\delta t = Tf_{RF}/\Delta$. Therefore, although the complete output signal at any one time does not represent a true distribution of the target positions corresponding to *that time*, that same signal does represent a true distribution of what the target positions *will be* Tf_{RF}/Δ seconds later. Thus, what might seem to be a complicated mixture of positive and negative range errors can be viewed simply as an "error of interpretation of the range data." (This time error is a constant of the radar system and can easily be handled by a computer.) One can hardly overemphasize the importance of such an extremely simple interpretation to unscramble the Doppler effect on the linear-FM Chirp signal!

The addition of shaping networks to reduce the side-lobe levels will not modify the above interpretation in any qualitative manner. The returning signal that has experienced a Doppler shift no longer passes through

the central symmetric region of the weighting network. By forcing the signal to pass through the filter away from the center, an over-all attenuation is introduced accompanied by some modification of the signal details (see Section 3.4.4). The principal feature of concern here is the inevitable signal attenuation and the time lead or lag caused by the receiver delay equalizer.

A useful pictorial representation of this state of affairs is afforded by the ambiguity diagrams of Woodward[10] (p. 118). Similar pictorial representations have been discussed earlier by Gabor[19] and Ville,[20] but the interpretation that will be stressed here was first emphasized by Woodward. Ambiguity diagrams are *joint*-response diagrams in both Doppler frequency and response time. They are most commonly calculated for the case of matched filters but need not be restricted to this ideal situation. Figs. 38(a), 38(b) and 38(c) illustrate, in a qualitative manner, ambiguity diagrams for (a) a long constant-frequency signal, (b) a short constant-frequency signal, and (c) a linear-FM Chirp signal. In each, the matched-filter characteristics, or nearly so, are assumed. The region of heavy shading schematically illustrates the region of strong signal response. The long constant-frequency signal, Fig. 38(a), possesses a good velocity resolution capability but a correspondingly poor range resolution capability. The pattern generated by the short signal shown in Fig. 38(b) has the opposite characteristics: good range resolution and poor velocity resolution. The interpretation to be placed on these ambiguity diagrams may now be readily understood. The shaded region — that is, the region of strong response — represents the "distribution of ambiguity" in interpreting the response of a single point target or in the ability to distinguish and identify several point targets separated either in range or velocity or both. Fig. 38(c) shows a qualitative ambiguity diagram for a linear-FM Chirp system. The Chirp scheme possesses resolution ambiguities situated along an *inclined* axis. The information contained in a *single* returning Chirp signal does not provide an unambiguous determination of both range and velocity, but represents only one quantity, which is a linear combination of these two variables. Unless the target velocity is known *a priori*, it is necessary to obtain several return signals so that a separate velocity determination is possible. For a large class of applications, this method represents an adequate solution to the ambiguity presented either by a Chirp signal or by a short constant-frequency signal.

The question naturally arises as to why one should be confined to ambiguity diagrams with the properties illustrated in Fig. 38. Why not use a signal leading to a small, concentrated shaded region that possesses

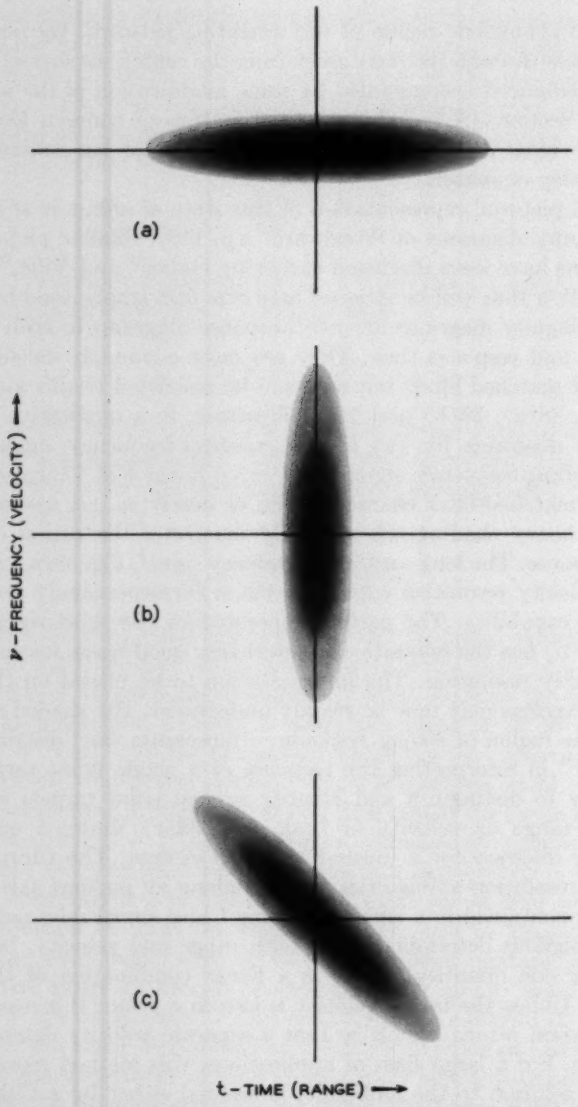


Fig. 38 — (a) Schematic ambiguity diagram for a long constant-frequency radar signal; the more shaded portions denote regions of greater signal response with the time, t, and for variously moving targets with Doppler frequency ν. (b) Schematic ambiguity diagram for a short constant-frequency radar signal; improved time resolution is secured at the expense of some discrimination against moving targets. (c) Schematic ambiguity diagram for a typical chirp signal; an approaching target with a positive Doppler shift has a response attenuated in amplitude and shifted to an earlier time.

simultaneous range and velocity resolution? Although this question is not directly related to the study of Chirp systems, it is a fascinating issue clearly encompassed by the *spirit* of the Chirp scheme. A companion paper[7] considers this question in more detail and investigates an important class of ambiguity diagrams.

IV. CONCLUSION

This paper has presented in detail the theory and design of Chirp radars. Throughout the paper the FM characteristic intentionally imposed on the transmitted signal took but one form: a uniform linear FM. Needless to say, such a simple specification as linear FM represents just one of numerous, more complex FM schemes that might be used. However, the simplicity of the Chirp radar would perhaps make it easier to construct than any of the hierarchy of alternative methods. This important feature of ease of construction increases in importance when it is realized that it is theoretically possible to design a high-performance linear-FM Chirp radar whose efficiency is only slightly below the ideal maximum; such is the result of the analysis presented in this paper.

In summarizing the contents of this paper the following remarks can be made:

i. Passive generation of the transmitted signal appears to be nearly as efficient as an ideal active generation process. In addition, passive generation is very desirable from a practical standpoint.

ii. The combination of receiver collapsing and weighting networks provides a good approximation to a matched filter. Moreover, the desirable signal property of very low side lobes is obtained by sacrificing only a few decibels in the signal-to-noise ratio.

iii. Even in the presence of substantial Doppler shifts the output signal maintains low side lobes. The most significant effect on the return signal is a time translation proportional to the velocity of the moving target. But by a simple data "reinterpretation" it is possible to nullify the effect of this time translation. This significant property follows only in the case of linear FM such as in Chirp.

iv. The design of Chirp radars with dispersion factors of 100 is quite reasonable; such radars are presently in the experimental stages of development. With a dispersion factor of 100, a threefold increase in range is to be expected.

V. ACKNOWLEDGMENTS

The authors are indebted to C. N. Nebel for his advice and sugges-

tions during the course of this work. They also take pleasure in acknowledging the assistance of the computing personnel and facilities of Bell Telephone Laboratories at both Whippany and Murray Hill, N. J.

REFERENCES

1. Darlington, S., U. S. Patent No. 2,678,997, May 18, 1954.
2. Hüttman, E., German Patent No. 768,068, March 22, 1940.
3. Sproule, D. O. and Hughes, A. J., British Patent No. 604,429, July 5, 1948.
4. Cauer, W., German Patent No. 892,772, December 19, 1950.
5. Dicke, R. H., U. S. Patent No. 2,624,876, January 6, 1953.
6. Cook, C. E., Proc. I.R.E., **48,** 1960, p. 310.
7. Klauder, J., this issue, p. 809.
8. MacColl, L. A., unpublished manuscript.
9. Taylor, T. T., I.R.E. Trans., **AP-3,** 1955, p. 316.
10. Woodward, P. M., *Probability and Information Theory, with Applications to Radar*, McGraw-Hill, New York, 1953.
11. Van Wijngaarden, A. and Scheen, W. L., *Tables of Fresnel Integrals*, N. V. Noord-Hollandsche Uitgevers Maatschappij, Amsterdam, 1949.
12. Campbell, G. A. and Foster, R. M., *Fourier Integrals for Practical Applications*, D. Van Nostrand Co., New York, 1942.
13. Wheeler, H. A., Proc. I.R.E., **27,** 1939, p. 359.
14. Burrows, C. R., Proc. I.R.E., **27,** 1939, p. 384.
15. Dolph, C. L., Proc. I.R.E., **34,** 1946, p. 335.
16. Kallman, H. E., Proc. I.R.E., **28,** 1940, p. 302.
17. Graham, R. S. and Sperry, R. V., U. S. Patent No. 2,760, 164.
18. Bellows, B. C. and Graham, R. S., B.S.T.J., **36,** 1957, p. 1429.
19. Gabor, D., J.I.E.E., **93,** Pt. III, 1946, p. 429.
20. Ville, J., Cables & Trans., **2,** 1948, p. 61.

The Design of Radar Signals Having Both High Range Resolution and High Velocity Resolution

By J. R. KLAUDER

(Manuscript received April 5, 1960)

Radar systems that yield simultaneous information regarding the range and velocity of a target would be useful in certain applications. A discussion is presented of some fundamental limitations on simultaneous range-velocity determination in terms of ambiguity diagrams, with the aid of a quantum mechanical analog to the Wigner distribution function. A sequence of signals is found in which the signals yield both range and velocity information with increasing accuracy. However, this desired property is not accompanied by a waveform suitable for maximum operating efficiency.

I. INTRODUCTION

The very large power requirements of present-day radars clearly demand full utilization of the potential of existing equipment. Unfortunately, an efficient equipment use does not always lead directly to a signal possessing desired high resolution characteristics. The "Chirp" scheme, which employs linear frequency modulation and is discussed in detail in the accompanying paper,[1] represents one method by which the desirable high resolution properties may be secured with an optimum equipment utilization. However, in Section 3.5 of the accompanying paper it was pointed out that the Chirp scheme, like ordinary radar techniques, possesses an inherent ambiguity in a simultaneous determination of both the range and velocity of a moving target. Ambiguity diagrams,[2] which are representations of joint response functions in both time and Doppler frequency, provide a pictorial representation of the uncertainty in determining range (time) and velocity (Doppler frequency). Fig. 1, (which is a duplicate of Fig. 38 of the accompanying paper) schematically illustrates the ambiguity diagrams corresponding to three different transmitted signals: (a) a long constant-frequency sig-

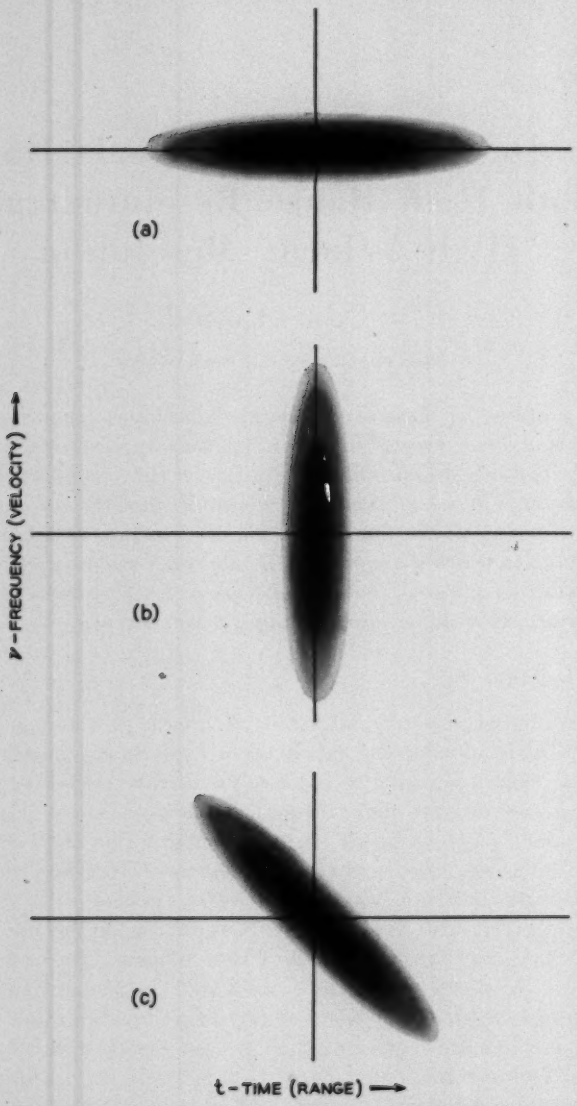


Fig. 1 — (a) Schematic ambiguity diagram for a long constant-frequency radar signal; the more shaded portions denote regions of greater signal response with time, t, and for variously moving targets with Doppler frequency, ν. (b) Schematic ambiguity diagram for a short constant-frequency radar signal; improved time resolution is secured at the expense of some discrimination against moving targets. (c) Schematic ambiguity diagram for a typical Chirp signal; an approaching target with a positive Doppler shift has a response attentuated in amplitude and shifted to an earlier time.

nal, (b) a very short constant-frequency signal and (c) a Chirp signal with linear frequency modulation.

For certain applications it may be desirable to transmit a signal whose ambiguity function is highly peaked only about the point $t = \nu = 0$. If such a signal exists, it would permit a high resolution in both range and velocity simultaneously. In Section II of this paper some properties of ambiguity functions are discussed with the aid of quantum mechanical analogs, and one interesting general property is discussed in detail: If a requirement of rotational invariance is imposed on ambiguity functions (a concept made precise in Section II) a sequence of signals is obtained for which the peaking of the ambiguity function is steadily increased. While the signals in this class achieve the desired behavior in the ambiguity function, the transmitter waveforms are not uniform in amplitude and, therefore, cannot be transmitted with the same efficiency as the uniform amplitude Chirp signal. The notation used in the following analysis is the same as that in the accompanying paper.

II. SOME GENERAL PROPERTIES OF AMBIGUITY DIAGRAMS

For convenience in the analysis, only a single target, which gives rise to a Doppler shift $\nu = vf_{RF}/c$ is studied, where v is the target velocity, f_{RF} is the radio-frequency carrier and c is the velocity of light. If the spectrum of an arbitrary complex transmitted signal is denoted by $\tilde{\epsilon}(f)$ then the received signal becomes $\tilde{\epsilon}(f - \nu)$. Only receiving networks that optimize the signal-to-noise ratio will be discussed. For an arbitrary complex signal $\epsilon(t)$, this optimum filter is the so-called matched filter for that signal. If $\tilde{\epsilon}(f)$ denotes the Fourier transform of the signal $\epsilon(t)$, then the matched filter $\tilde{Y}_m(f)$ may be defined[1] as $\tilde{\epsilon}^*(f)$. With a matched filter in the receiver, the response spectrum is determined by $\tilde{\epsilon}^*(f)\tilde{\epsilon}(f - \nu)$. The Fourier transform determines the time response:

$$\chi(t,\nu) = \int_{-\infty}^{\infty} \tilde{\epsilon}^*(f)\tilde{\epsilon}(f - \nu)e^{2\pi ift}\, df, \tag{1}$$

which is written as a two-parameter function. This is the expression given by Woodward,[2] and can be used perfectly well to specify the ambiguity diagrams. However, for the analogy to be drawn to single-particle nonrelativistic quantum mechanics, it is convenient to develop an alternate expression for the ambiguity function amplitude.

Suppose that the transmitter spectrum, $\tilde{\epsilon}(f)$, is concentrated about some high RF frequency, so that

$$\tilde{\epsilon}(f) = \tilde{\mathcal{E}}(f - f_{RF}), \tag{2a}$$

where $\tilde{\mathcal{E}}(f)$ is a spectrum centered about zero frequency. It follows from (2a) that

$$\epsilon(t) = \mathcal{E}(t)e^{2\pi i f_{RF} t}. \tag{2b}$$

Therefore, $\mathcal{E}(t)$ represents the generalized envelope of the transmitted signal. If (2a) is substituted into (1), one readily finds that

$$\chi(t,\nu) = e^{2\pi i (f_{RF}+\nu/2)t} \int_{-\infty}^{\infty} \tilde{\mathcal{E}}^* \left(f + \frac{\nu}{2}\right) \tilde{\mathcal{E}} \left(f - \frac{\nu}{2}\right) e^{2\pi i f t}\, df, \tag{3a}$$

or, when expressed in terms of $\mathcal{E}(t)$,

$$\chi(t,\nu) = e^{2\pi i (f_{RF}+\nu/2)t} \int_{-\infty}^{\infty} \mathcal{E}^* \left(\tau - \frac{t}{2}\right) \mathcal{E} \left(\tau + \frac{t}{2}\right) e^{2\pi i \nu \tau}\, d\tau. \tag{3b}$$

Let the carrier in the presence of a Doppler shift be taken as $f_{RF} + \nu/2$. With this choice of carrier,

$$\rho(t,\Omega) = \int_{-\infty}^{\infty} \mathcal{E}^* \left(\tau - \frac{t}{2}\right) \mathcal{E} \left(\tau + \frac{t}{2}\right) e^{i\Omega\tau}\, d\tau \tag{4}$$

is defined to be the complex two-parameter response envelope. For convenience, the Doppler angular frequency $\Omega = 2\pi\nu$ has been introduced in (4). It is the absolute value of $\rho(t,\Omega)$, which represents the detected output. The remaining analysis uses the definition made in (4); it is also the form studied by Gabor[3] and Ville.[4]

No loss of generality is made by assuming the transmitted signal carries unit energy:

$$\int_{-\infty}^{\infty} |\epsilon(t)|^2\, dt = \int_{-\infty}^{\infty} |\mathcal{E}(t)|^2\, dt = \int_{-\infty}^{\infty} |\tilde{\mathcal{E}}(f)|^2\, df = 1. \tag{5}$$

With this normalization, it follows from (4) that

$$\rho(0,0) = \int_{-\infty}^{\infty} |\mathcal{E}(\tau)|^2\, d\tau = 1. \tag{6}$$

This result represents a constraint that requires a portion of the "ambiguity" to remain in the vicinity of the point $t = \Omega = 0$. By a straightforward integration using (4) and (5), the following result may be established:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\rho(t,\Omega)|^2\, dt\, d\Omega = 1. \tag{7}$$

[This result applies equally well to a similar integration involving $\chi(t,\nu)$.] The interpretation given to (7) is as follows: There is, quali-

tatively speaking, a total ambiguity equal to "one", which must be distributed throughout the t,Ω plane. The approximately equal amounts of shaded area in Fig. 1 are intended to be a reflection of this constraint. This result seems to preclude any simple solution to achieve simultaneously an arbitrarily high resolution in both range and velocity.

Another observation worth making about the three diagrams in Fig. 1 is that it appears they could be obtained from one another by a "rotation" about the origin of the t,Ω plane. It will be demonstrated below that this is a general result. That is, for every ambiguity diagram defined by $\rho(t,\Omega)$ there exists another signal that generates another diagram differing from the first only by a rotation about $t = \Omega = 0$. The use of the word "rotation" needs some explanation here. Actually, the rotation can proceed along arbitrary elliptical contours. However, for convenience, it is assumed that the scales of time and Doppler frequency have been chosen so that the rotation is nearly circular.

The rotation theorem, as well as others like it, is most readily established by using methods common in the study of quantum mechanics. One of the reasons for choosing the form of (4) in preference to that of (1) was to facilitate the comparison with a quantity studied previously in quantum mechanics. Only a brief sketch is presented here of the comparison used to obtain most of the remaining results of this section.

The quantity of interest in quantum mechanics is the so-called Wigner[5] distribution function defined by

$$P(q,p) = \frac{1}{h}\int_{-\infty}^{\infty} \psi^*\left(q - \frac{y}{2}\right) e^{-iyp/\hbar} \psi\left(q + \frac{y}{2}\right) dy. \tag{8}$$

Here, $h = 2\pi\hbar$ is Planck's constant and $\psi(y)$ represents the Schrödinger wave function for a single particle in one dimension. Of more direct interest is the characteristic function for this distribution, i.e. its double Fourier transformation:

$$\begin{aligned} \rho(y,\kappa) &= \int P(q,p) e^{i(\kappa q + yp)/\hbar}\, dq\, dp, \\ &= \int \psi^*\left(q - \frac{y}{2}\right) e^{i\kappa q/\hbar} \psi\left(q + \frac{y}{2}\right) dq. \end{aligned} \tag{9}$$

If this result is compared with (4), one notes an exact functional equivalence, with t playing the role of y and Ω playing the role of $\kappa/\hbar$. Equation (9) is perhaps easier than (8) to study, and has been treated in some detail by Bass,[6] Moyal[7] and Groenewold;[8] further references may be found in these works. The relative simplicity of (9) stems from its simple representation in the quantum mechanical Hilbert space:

$$\rho(y,\kappa) = (\Psi, e^{i(\kappa\mathbf{q}+\nu\mathbf{p})/\hbar}\Psi). \tag{10}$$

Here Ψ denotes an arbitrary normalized vector and $\mathbf{q}$ and $\mathbf{p}$ are the usual operators satisfying $\mathbf{qp} - \mathbf{pq} = i\hbar$. Numerous additional relations may be found, especially in the work of Moyal.

The necessary signal transformation to generate a rotated ambiguity diagram will not be derived; the result will be stated without proof. Suppose that $\mathcal{E}(t)$ is an envelope which generates a particular ambiguity function $\rho(t,\Omega)$ according to (4). The modified envelope, $\mathcal{E}_\theta(t)$, must be chosen so that its ambiguity function is defined by

$$\rho_\theta(t,\nu) = \rho\left(t\cos\theta + \frac{\nu}{\gamma}\sin\theta,\ \nu\cos\theta - \gamma t\sin\theta\right). \tag{11}$$

Here, γ represents the ratio of the "angular frequency" axis to the "time" axis of the ellipse of rotation and θ represents the "angle" of rotation. When $\gamma \approx 1$, rotation is along nearly circular orbits. The solution for $\mathcal{E}_\theta(t)$ is given by

$$\mathcal{E}_\theta(t) = \int_{-\infty}^{\infty} U(t,\tau;\theta)\mathcal{E}(\tau)\,d\tau, \tag{12}$$

where the kernel $U(t,\tau;\theta)$ is expressed as follows:

$$U(t,\tau;\theta) = \sqrt{\frac{\gamma}{2\pi i\sin\theta}}\exp\left\{i\gamma\left[\left(\frac{t^2+\tau^2}{2}\right)\cot\theta - t\tau\csc\theta\right]\right\}. \tag{13}$$

Equation (13) follows from the properties of the well-known harmonic oscillator problem, and is just one of the relations readily obtained with the techniques of quantum mechanics.

The particular relations expressed in (12) and (13) are perhaps not as interesting as an additional result that can be derived from them. In the discussion of the ambiguities present in the interpretation of the diagrams in Fig. 1 it seemed that the resolution was relatively good in one direction but was poor in the perpendicular direction. It is just such a variation with direction that is *not* wanted in attempting to achieve a signal whose ambiguity function is concentrated about the point $t = \Omega = 0$. For clarity, assume again that the t and Ω axes have already been scaled for the specific system under study. Then, to secure (as nearly as possible) an ambiguity diagram concentrated about the origin, it appears desirable to seek a *rotational invariance of the ambiguity function*; i.e., $\rho_\theta(t,\Omega) = \rho(t,\Omega)$. It is a simple matter to demonstrate that $\rho_\theta(t,\Omega) = \rho(t,\Omega)$ if and only if

$$\mathcal{E}_\theta(t) = \int_{-\infty}^{\infty} U(t,\tau;\theta)\mathcal{E}(\tau)\,d\tau = e^{-i\lambda\theta}\mathcal{E}(t), \tag{14}$$

where λ represents a suitable constant. There are a denumerable number of solutions to (14), a typical one being denoted by $\mathcal{E}_n(t)$, where $\lambda_n = n + \frac{1}{2}$. These solutions, familiar as the eigenfunctions of the harmonic oscillator, are defined by the following relation:

$$\mathcal{E}_n(t) = \frac{\gamma^{\frac{1}{4}}}{\sqrt{\pi^{\frac{1}{2}} n! 2^n}} H_n(\sqrt{\gamma} t) e^{-\gamma t^2/2}, \tag{15}$$

where $H_n(z)$ represents the nth Hermite polynomial defined, for example, by

$$\mathrm{H}_n(z) = (-1)^n e^{z^2} \frac{d^n}{dz^n} e^{-z^2}. \tag{16}$$

Figs. 2(a), 3(a), 4(a) and 5(a) show four curves of $\mathcal{E}_n(t)$ for $n = 0$, 4, 10 and 26, respectively; each curve is symmetric about $t = 0$ and is only shown for positive values of t. Note also a change in scale in the different figures. In each example γ was chosen equal to one.

It is a well-known fact that the functions $\mathcal{E}_n(t)$ defined by (15) have spectra, $\tilde{\mathcal{E}}_n(f)$, which are specified by the very same functional form.† Consequently, Figs. 2(a) through 5(a) also represent the frequency spectra for the same n values. One can see from these four figures that, as n increases, both the pulse width and bandwidth increase. The pulse-width-bandwidth product increases as $(n + \frac{1}{2})^2$ as n is increased. One may hope, therefore, that the ambiguity function becomes more nearly concentrated about the origin as n increases. The normalization condition (7) must still be satisfied, but, for large n, the remaining "ambiguity" is thinly spread out over a large region of the t,Ω plane.

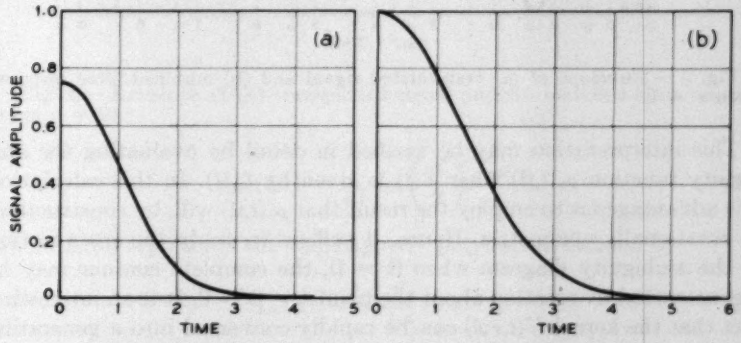


Fig. 2 — (a) Envelope of the transmitted signal when the parameter $n = 0$; in this case the curve is gaussian. (b) Envelope of the "rotationally invariant" matched-filter response when $n = 0$; here the curve is also gaussian.

† See Pair 702 in the tables of Campbell and Foster.[9]

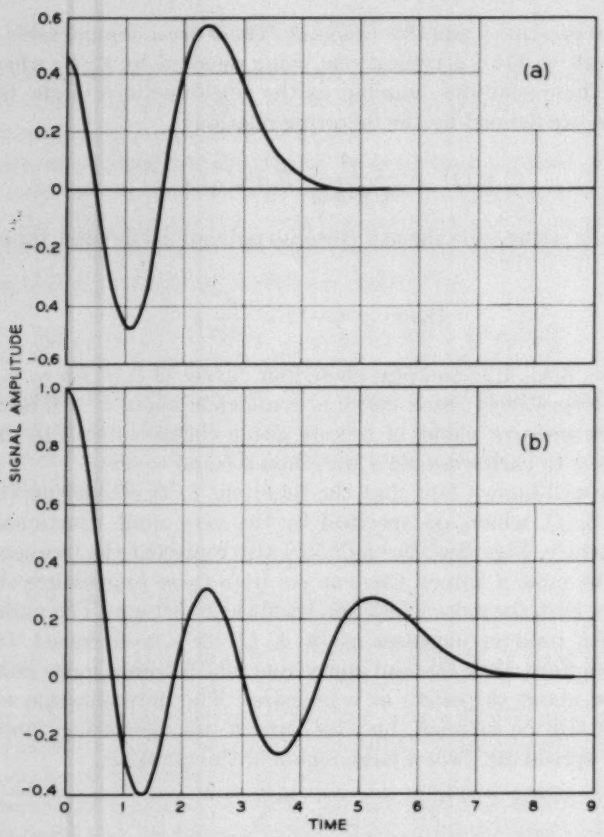


Fig. 3 — Envelope of (a) transmitted signal and (b) matched-filter response when $n = 4$.

This interpretation may be verified in detail by evaluating the ambiguity function $\rho_n(t,\Omega)$ when $\mathcal{E}(t)$ is given by $\mathcal{E}_n(t)$. In this calculation it is advantageous to employ the result that $\rho_n(t,\Omega)$ will, by construction, be rotationally symmetric. Hence, it suffices to study the cross section of the ambiguity diagram when $\Omega = 0$; the complete contour may be reconstructed by rotation about the point $t = \Omega = 0$. It is an interesting fact that the kernel $U(t,\tau;\theta)$ can be rapidly converted into a generating function for the various response signals $\rho_n(t,0)$. The details of such a calculation are essentially present in the work of Moyal.[7] Either from the generating function or by a direct use of the definition for $\rho(t,\Omega)$ one finds the zero Doppler response to be

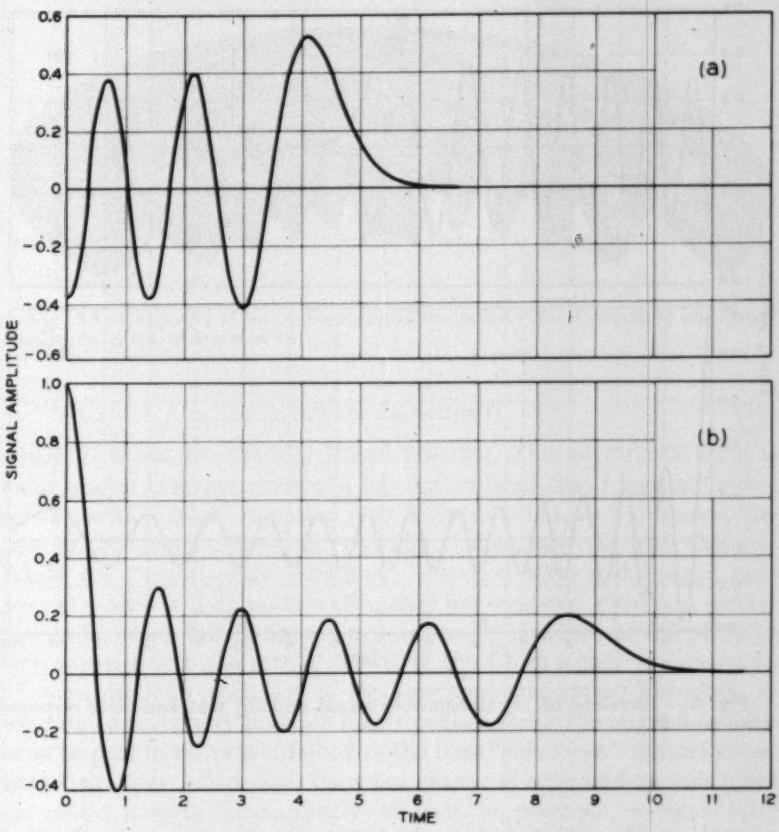


Fig. 4 — Envelope of (a) transmitted signal and (b) matched-filter response when $n = 10$.

$$\rho_n(t,0) = \frac{1}{n!} L_n\left(\frac{\gamma t^2}{2}\right) e^{-\gamma t^2/4}, \tag{17}$$

where $L_n(z)$ represents the nth Laguerre polynomial defined by

$$L_n(z) = e^z \frac{d^n}{dz^n}(z^n e^{-z}). \tag{18}$$

The output signal $\rho_n(t,0)$ is illustrated in Figs. 2(b), 3(b), 4(b) and 5(b) for $n = 0$, 4, 10 and 26, respectively. Since these functions are clearly symmetric only positive t values have been included (note the change

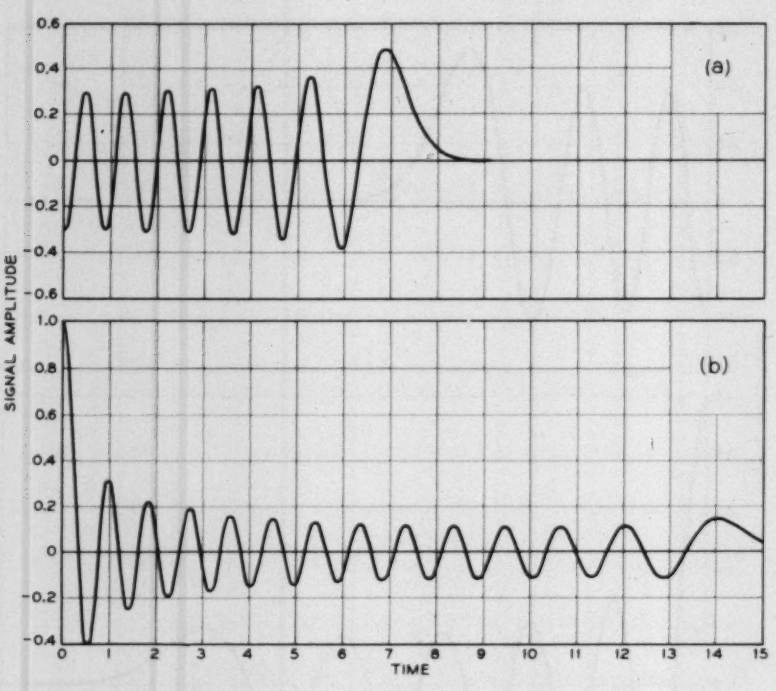


Fig. 5 — Envelope of (a) transmitted signal and (b) matched-filter response when $n = 26$.

of scale in the different figures). Again, these curves were obtained assuming $\gamma = 1$. In order to illustrate the rotationally invariant character of these signals, Fig. 6 shows a cut-away of a three-dimensional projected view of the complete ambiguity amplitude $\rho(t,\Omega)$ corresponding to the case $n = 10$.

Fig. 6 illustrates the tendency of the ambiguity function to peak about the origin of the t,Ω plane. Furthermore, it demonstrates that a narrow peak is achieved only when the remaining "ambiguity" is spread out over a relatively large domain of the t,Ω plane.

If input and response envelopes ($\mathcal{E}_n(t)$ and $\rho_n(t,\Omega)$) with higher values of n are considered, the ambiguity function becomes more sharply peaked and more spread out. A study of the behavior of the zero Doppler response for large values of n shows that

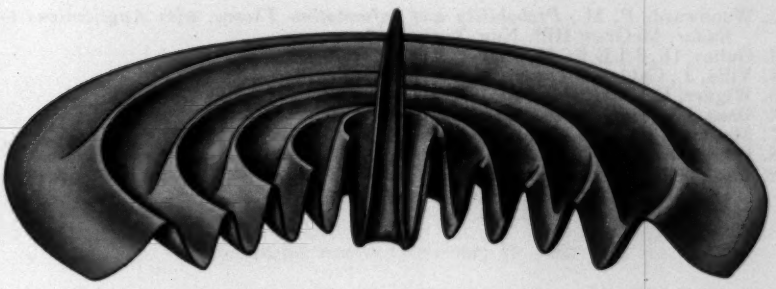


Fig. 6 — Cut-away of three-dimensional projected view of complete ambiguity amplitude $\rho(t,\Omega)$ when $n = 10$.

$$\rho_n(t,0) \sim J_0(\sqrt{2\gamma n}\, t), \tag{19}$$

where J_0 is the zeroth-order Bessel function. This asymptotic form is valid when t is greater than $\sqrt{n}$ but not so large that t becomes comparable with n itself. Equation (19) illustrates that, as n increases, the *form* of the response does not markedly change, only the *scale* changes. While the "rotationally invariant" response signals discussed here possess a narrow 3-db pulse width, they are endowed with long oscillatory tails, which fall off approximately as t^{-4}. This rate of fall-off is to be compared with the rate of fall-off of the Chirp signal — namely, as t^{-1}, which follows from (19) of the accompanying paper. Therefore, in securing an ambiguity function highly peaked about the origin a penalty must be paid in the rate of fall-off of the time "side lobes." Nevertheless, as the parameter n increases there is a change of scale and the side lobes are pulled inward. Consequently, at least in principle, a signal $\mathcal{E}_n(t)$ could be chosen that maintained side-lobe levels below an arbitrary level outside a prescribed ellipse in the t,Ω plane. It is interesting to speculate whether there exists some signal envelope $\mathcal{E}(t)$ that could be efficiently transmitted in addition to having a highly peaked ambiguity function. Unfortunately, this question remains unanswered.

III. ACKNOWLEDGMENTS

It is a pleasure to thank Miss B. Cetlin for her assistance in the computations required in this work.

REFERENCES

1. Klauder, J. R., Price, A. C., Darlington, S. and Albersheim, W. J., this issue, p. 745.

2. Woodward, P. M., *Probability and Information Theory, with Applications to Radar*, McGraw-Hill, New York, 1953.
3. Gabor, D., J.I.E.E., **93**, Pt. III, 1946, p. 429.
4. Ville, J., Cables & Trans., **2**, 1948, p. 61.
5. Wigner, E., Phys. Rev., **40**, 1932, p. 749.
6. Bass, J., Compt. rend., **221**, 1945, p. 46.
7. Moyal, J. E., Proc. Camb. Phil. Soc., **45**, 1949, p. 99.
8. Groenewold, H. J., Physica, **12**, 1946, p. 405.
9. Campbell, G. A. and Foster, R. M., *Fourier Integrals for Practical Applications*, D. Van Nostrand Co., New York, 1942.

The TJ Radio Relay System

By J. GAMMIE and S. D. HATHAWAY

(Manuscript received February 19, 1960)

The TJ radio relay system is a broadband microwave facility that operates in the 10,700- to 11,700-mc common carrier frequency band. It has been specifically designed for short-haul transmission of either multichannel telephone or television circuits. Transmission performance and the over-all system description are presented, as well as some early field applications.

TABLE OF CONTENTS

I. Introduction 821
II. Objectives 822
2.1 Area of Application 822
2.2 Development Objectives and Performance Characteristics 823
2.3 Transmission Objectives 824
III. Transmission Plan 825
IV. Description of System 828
4.1 General Description 828
4.2 Diversity Switching Arrangements 845
4.3 Nondiversity Applications 850
4.4 Order-Wire, Alarm and Control System 850
4.5 Antenna Systems 853
4.6 Connecting Circuits 859
4.7 Standby Power 861
V. Equipment Features 862
5.1 Transmitter-Receiver Bay 862
5.2 Auxiliary Bay 865
5.3 Interconnecting Arrangements 867
VI. System Maintenance and Test Equipment 870
VII. Applications of TJ Radio Systems 871
7.1 Path Selection 871
7.2 Typical TJ Installations 874
VIII. Acknowledgment 877
References 877

I. INTRODUCTION

During the past decade, microwave radio in the Bell System has had a phenomenal growth. In terms of route mileage and number of circuits, this growth has been primarily in the long-haul field in the 4,000-mc common carrier band. The TD-2 system[1,2] now criss-crosses the continent several times and provides facilities for telephone and

television to almost every part of the continental United States.* In many areas it is already loaded to capacity and, hence, its use for other than backbone service is becoming increasingly restricted. In the next decade it is expected that there will be a large demand for short-haul microwave facilities along backbone routes and remote rural areas. These will have to be supplied by systems operating in common carrier bands other than the 4,000-mc band. To allow for orderly growth of the Bell System radio plant, the 11,000-mc band has been selected for short-haul service needs where the maximum channel cross section might be only a few hundred telephone circuits.

In the past, the telephone companies have used 4,000-mc TE equipment,[3] secondary TD-2 arrangements or other currently available 6,000-mc common carrier equipment to fulfill their short-haul needs. Feasibility studies[4] by the radio research group at Bell Telephone Laboratories and systems engineering studies made in cooperation with the telephone companies indicated the possibility of and need for a new economical short-haul system that would permit the dropping and adding of circuits at each repeater or alternatively, be capable of transmitting monochrome or NTSC color television. Because of the potential interference problems with the 4,000-mc TD-2 system, and the 6,000-mc TH system[5] now being installed in some sections of the country, the new TJ system has been developed for the 10,700–11,700-mc common carrier band.

II. OBJECTIVES

2.1 *Area of Application*

There are a great number of uses for a flexible, economical short-haul radio system such as TJ. A partial list of these applications includes:

(a) relief on open-wire or cable routes now at full capacity;

(b) added facilities along open-wire and secondary cable routes for improved reliability;

(c) television side-legs and short-haul message facilities branching off backbone routes;

(d) short-haul message facilities on existing backbone TD-2 or TH radio routes;

* At present, more than 15 million (27 per cent) of the long distance telephone circuit miles and more than 60,000 (78 per cent) of the intercity television circuit miles of the Bell System are provided by microwave radio relay.

(e) new facilities to locations where wire construction is difficult;
(f) alarm, control and order-wire facilities on backbone radio routes;
(g) general purpose local television service;
(h) bypass service in large city areas where the TD-2 system is at capacity.

2.2 *Development Objectives and Performance Characteristics*

Prior to the beginning of the development program, a set of objectives was specified reflecting the best judgment at that time as to the features and capabilities necessary for the new system. These objectives have been reviewed and modified from time to time as the development of the new system progressed and, in most instances, the original requirements for the TJ system have been met.

2.2.1 *Frequency Band and Allocation Plan*

The TJ system operates in the 11,000-mc common carrier frequency band and provides either message or television service for end-link or short-haul applications. A frequency allocation plan has been devised to permit operation of six two-way TJ channels on the same route with common antennas for transmitting and receiving. One-for-one frequency diversity, with a simple automatic switch has been provided on an optional basis, and little or no cost penalty has been incurred by systems not having such switches.

2.2.2 *Telephone and Television Capacity*

The TJ system has been engineered to transmit 96 channels of ON-2 multiplex[6] or 240 channels of L multiplex through nine repeaters for a total distance of 200 to 300 miles. The system is also being engineered to transmit one monochrome or NTSC color television signal, meeting end-link and short-haul objectives over distances of approximately 100 miles on each radio channel. Sufficient video bandwidth has been provided to transmit the audio portion of the television program along with the video signal on a multiplex basis in the vicinity of 6 mc.

2.2.3 *Order Wire and Alarm*

An order-wire and alarm facility has been provided which will work over the radio for TJ telephone systems or over a wire facility for one-way television routes.

2.2.4 *Stand-by Power*

The stand-by power equipment provides the minimum discontinuity consistent with system cost objective and with other sources of system failure.

2.2.5 *Economics*

One of the primary objectives of this development has been to provide system arrangements and operating features at the lowest first cost and annual charge consistent with meeting Bell System requirements. Equipment arrangements have been designed to minimize job engineering and installation expense. Proper packaging has received careful consideration. Ease of maintenance is of prime importance in its relation to annual charges, and equipment arrangements have been devised with this in view.

2.3 *Transmission Objectives*

2.3.1 *General*

The TJ hops should be engineered to have a 50-db rms carrier-to-noise ratio during periods of free space transmission. The 50-db ratio is necessary to provide adequate margin over first circuit noise during periods of signal attenuation caused by rainfall. Propagation tests[7] conducted at 11,000 mc indicated that during periods of heavy rainfall the attenuation may be in the order of 40 db or more, depending upon the length and location of the radio path. Reliable protection against selective fades and equipment failure outages can be obtained with frequency diversity. Protection against rain attenuation, however, can only be assured by engineering sufficient fading margin into the system and by using path lengths appropriate to the particular area of the country.

2.3.2 *Telephone*

The TJ telephone end-link objective for cross-modulation and noise has been set at 32 dba at the 0 db transmission level point. This figure was arrived at by assuming that the random addition (power addition) of four such links to a backbone system should not degrade by more than 3 db the long-haul, heavy-route system objective of 38 dba. In practice, the degradation is not expected to exceed more than about 1 db in most cases because of shorter end-links and switching losses.

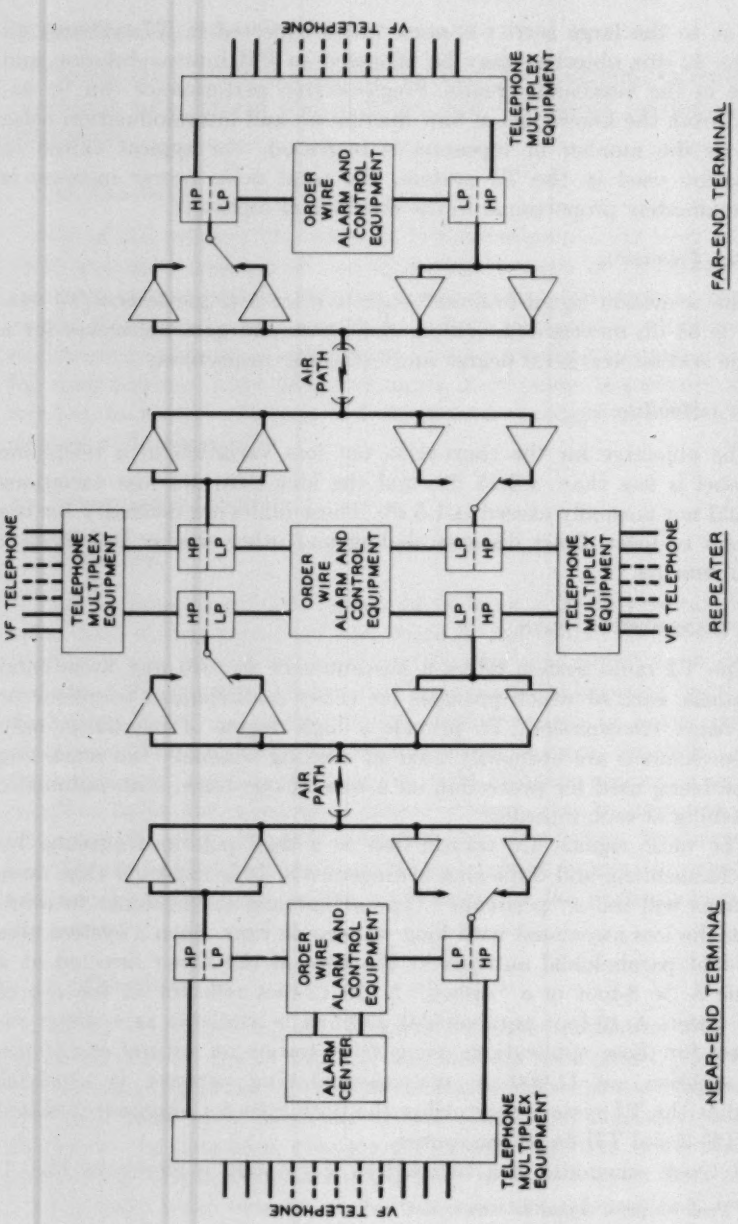


Fig. 1 — Block schematic of a two-section TJ system.

Due to the large carrier-to-noise ratios expected in TJ systems, all of the 32 dba objective may be allocated to FM intermodulation and noise in the baseband circuits. Single-section performance can be derived from the knowledge of how fluctuation and intermodulation noise add as the number of repeaters is increased. For typical values of deviation used in the TJ system, the total noise power increase is approximately proportional to the numbers of repeaters.

2.3.3 *Television*

The television signal-to-noise* objective for a single-section TJ system is 54 db unweighted. Differential phase and gain objectives for a single section are ±1.0 degree and ±0.5 db, respectively.

2.3.4 *Stability*

The objective for the short-term net loss variation in a telephone channel is less than ±0.25 db, and the long-term net loss variations should not normally exceed ±1.5 db. These limits are necessary for the system to meet direct distance dialing and other similar Bell System requirements.

III. TRANSMISSION PLAN

The TJ radio system offers a maximum of six two-way broadband channels, each of which provides for either multichannel telephone or television transmission. To provide a high degree of reliability, only three channels are ordinarily used as working channels, the remaining three being used for protection on a one-for-one basis, with automatic switching at each repeater.

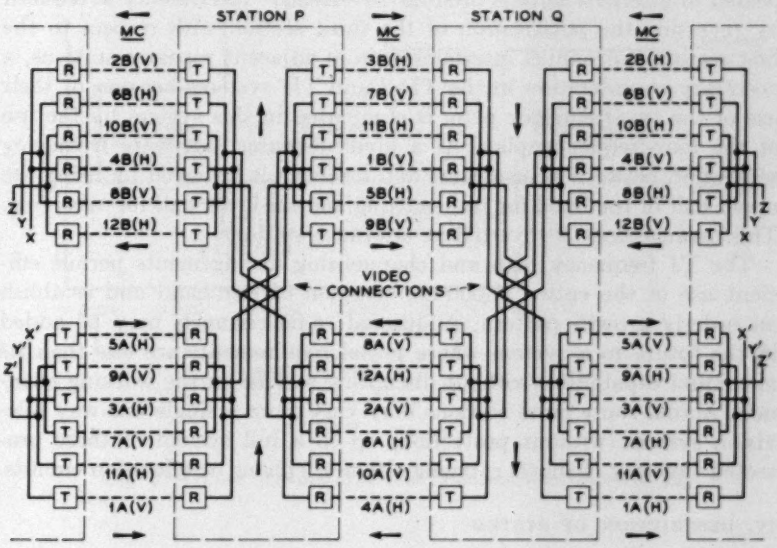
The radio signals are transmitted to a dual polarized antenna by RF channelizing and duplexing arrangements. It is expected that most systems will use a "periscope" type of antenna arrangement to minimize the loss associated with long waveguide runs. Such a system uses a 5-foot paraboloidal antenna at the base of the tower directed at a plane 6- × 8-foot or a "dished" 8- × 12-foot reflector at the top of the tower. A 10-foot paraboloidal antenna is available as a direct radiator for those applications using short towers on natural elevations. In addition, an 11,000-mc systems-combining network is available so that the TJ system may utilize the horn-reflector antennas installed on TD-2 and TH backbone routes.

A block schematic of a two-section TJ system is shown in Fig. 1.

* Peak-to-peak signal to rms noise.

The multiplex and control signals are combined by the high-pass–low-pass filter and feed the transmitters of the working and spare radio channel. At the receiving end, the selected radio receiver feeds into a similar filter combination, which separates the multiplex from the control signals. A 2600-cps pilot is transmitted over the system and is used to send alarms and orders between the various radio locations.

The TJ frequency plan is shown in Fig. 2. Because of the expected use of the "periscope" antenna system, the plan is based on the use of four frequencies for each two-way radio channel. The 10,700- to



Channel Number	Transmitter Frequency, kmc	Beat Oscillator Frequency, kmc	Channel Number	Transmitter Frequency, kmc	Beat Oscillator Frequency, kmc
4A	10.715	10.785	9B	11.245	11.315
1A	10.755	10.825	12B	11.285	11.355
10A	10.795	10.865	5B	11.325	11.395
11A	10.835	10.905	8B	11.365	11.435
6A	10.875	10.945	1B	11.405	11.475
7A	10.915	10.985	4B	11.445	11.515
2A	10.955	10.885	11B	11.485	11.415
3A	10.995	10.925	10B	11.525	11.455
12A	11.035	10.965	7B	11.565	11.495
9A	11.075	11.005	6B	11.605	11.535
8A	11.115	10.045	3B	11.645	11.575
5A	11.155	11.085	2B	11.685	11.615

Fig. 2 — TJ frequency assignment plan.

11,700-mc common carrier band is divided into 24 channels, each about 40 mc wide. In a given repeater section, only 12 of these are used, resulting in 80-mc spacing between midchannel frequencies. These channels are further divided into two groups of six for transmission in each direction. The polarization of the channels alternates between vertical and horizontal to provide 160-mc separation between signals having the same polarization, thereby substantially easing requirements on the channel-separation networks. The remaining 12 channel assignments are used in adjacent repeater sections. These frequencies are repeated in alternate hops. Potential "overreach" interference is reduced by reversing the polarization of the third section with respect to the first section. Cochannel interference from adjacent repeater stations, a necessary consideration in the TD-2 and TH systems because of their use of the two-frequency plan, is eliminated in this system by the use of the four-frequency plan. At a given repeater, adequate frequency separation between transmitters and receivers is achieved by using the upper half of the band for transmitting and the lower half for receiving. This arrangement is inverted at alternate stations.

The TJ frequency plan and channelizing arrangements permit efficient use of the entire 11,000-mc common carrier band and establish an orderly growth pattern. Additional radio channels may be added in the future to a system whose initial requirements are less than its maximum capabilities without disrupting service on the working channels. Actual route cross sections may vary from a single one-way television system, without protection, up to a full system of three protected two-way channels carrying either telephone or television circuits.

IV. DESCRIPTION OF SYSTEM

4.1 *General Description*

A basic building block in the TJ system is the transmitter-receiver bay. It consists of a frequency-modulated transmitter, a heterodyne-type receiver and regulated rectifiers operating from standard ac line voltages. A block schematic of the bay is shown in Fig. 3. An incoming microwave signal from the antenna system is selected by a channel-separation network of the type shown in Fig. 4. This network drops the desired channel and permits the remaining channels to pass through essentially unattenuated for selection in similar networks on adjacent bays. The selected channel is fed to the receiving modulator, which is preceded by a bandpass filter providing additional preselection. In the

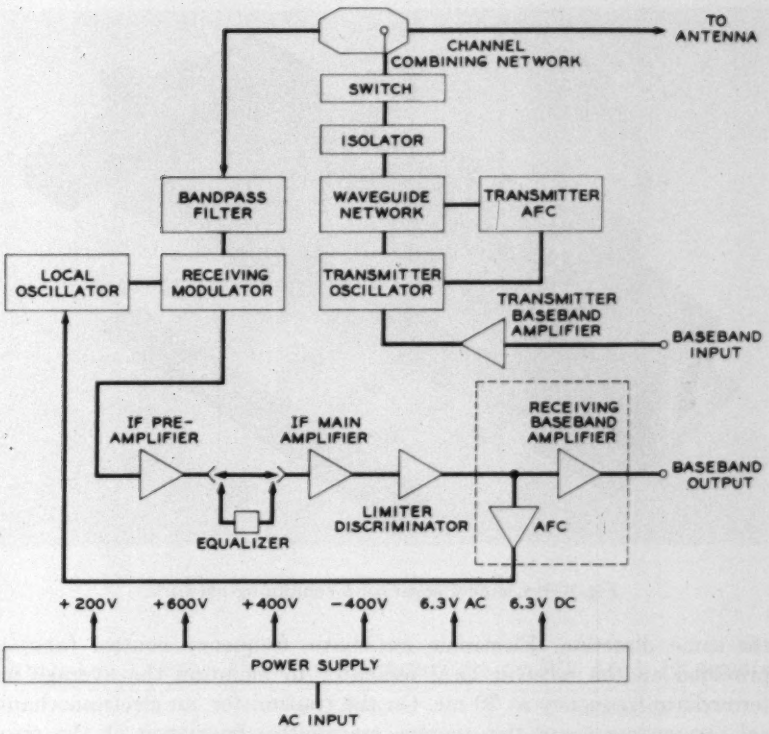


Fig. 3 — Block schematic of the TJ transmitter-receiver bay.

receiving modulator the signal is heterodyned with the output of a local oscillator to produce a difference or intermediate frequency (IF) of 70 mc. The IF signal is amplified and detected in the receiver to provide the original baseband intelligence, which may be applied either to the next transmitter or delivered to appropriate terminal equipment. In the transmitter, the baseband signal is amplified and applied to the repeller electrode of the transmitting klystron. The frequency-modulated output of the klystron is combined with outputs from adjacent transmitters in a channel-combining network similar to the receiver separation network. The transmitter outputs are then connected to the antenna system which in general will be simultaneously receiving from

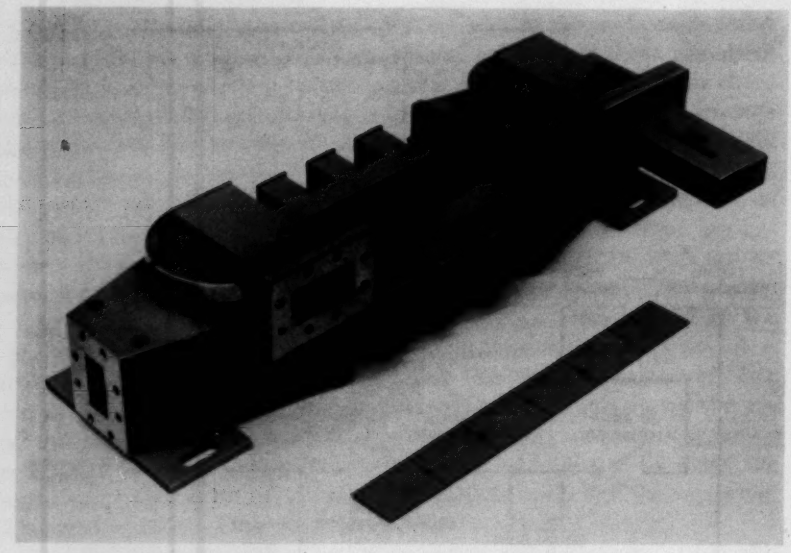


Fig. 4 — Channel separation-combining network.

the same direction. Electronic automatic frequency control (AFC) is provided on the receiver local oscillator to maintain the average intermediate frequency at 70 mc. On the transmitter, an electromechanical AFC system keeps the average transmitter frequency at the resonant frequency of a highly stable reference cavity.

The baseband type of repeater distinguishes the TJ system from the more common situation in long-haul microwave systems such as TD-2 and TH. In these, the signal remains in the frequency-modulated form throughout the amplification process at repeaters, and the baseband signal is only recovered through the use of special terminal equipment.

4.1.1 *Transmitter Radio Frequency Units*

The radio frequency output from the transmitter is provided by a Western Electric 445A klystron developed specifically for the TJ system. The same tube, illustrated in Fig. 5, is used as the receiver local oscillator. Typical operating characteristics in both applications are summarized in Table I.

Although the nominal output of the transmitter klystron is 0.5 watt when operated in the $2\frac{3}{4}$ mode, the actual output is dependent to

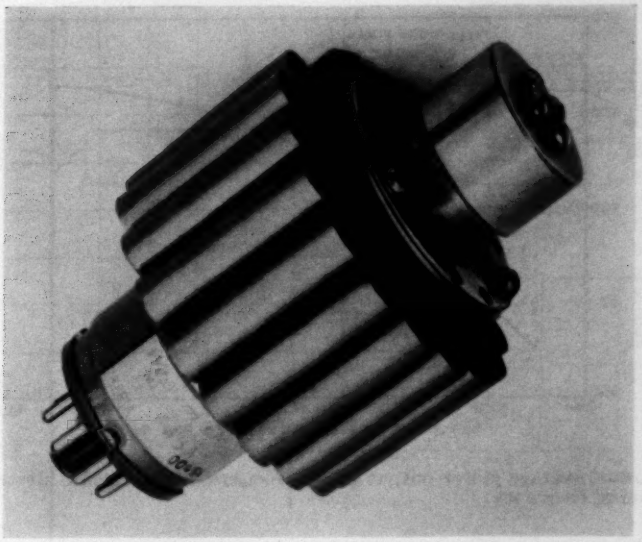


Fig. 5 — Western Electric 445A klystron.

TABLE I — TYPICAL OPERATING CONDITIONS OF WE 445A KLYSTRON

	Transmitter	Beat Oscillator
Resonator voltage	+600 volts	+400 volts
Resonator current	65 milliamperes	40 milliamperes
Repeller voltage	−250 volts	−125 volts
RF power output	400 milliwatts (minimum)	50 milliwatts (minimum)
Oscillating mode	$2\frac{1}{4}$	$3\frac{1}{4}$
Electronic tuning range	50 mc (minimum)	50 mc (minimum)
Repeller modulation sensitivity	0.8 mc/volt (minimum)	1.5 mc/volt (minimum)
Cooling	forced air	natural convection
Heater voltage	6.3 volts	
Heater current	0.9 ampere	
Repeller capacity	6 mmf (maximum)	
Mechanical tuning sensitivity	0.5 mc/angular degree	
Output	matched to WR90 waveguide	

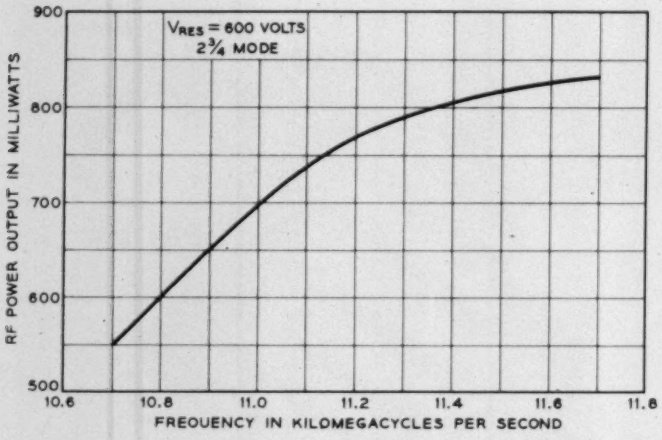


Fig. 6 — Average power output of the Western Electric 445A klystron as a function of frequency.

some extent on frequency. This dependence is demonstrated in Fig. 6, in which the average output of a number of production tubes is plotted against frequency.

Precautions against tube damage by positive repeller voltage have been included in both the transmitter and local oscillator circuits in the form of a clamping diode between repeller and cathode.

To reduce the maximum dc voltages on the bay and, hence, to simplify protective arrangements, the klystron body (resonator) operates at 600 volts above ground. An insulator between the tube output and its mating flange keeps this dc potential off the connecting waveguide.

The output from the transmitting klystron feeds a waveguide network, which serves the dual purpose of an AFC discriminator and power monitor. A schematic of the network is shown in Fig. 7. The first directional coupler feeds a waveguide hybrid, which, in conjunction with an invar reference cavity and a pair of silicon diodes, forms an RF discriminator. The operating principles of the discriminator have been described by Pound,[8] and a typical output characteristic is shown on the schematic. Zero output at the "tails" of the discriminator characteristic is controlled by the balance control, while the crossover point can be set to any frequency in the TJ band by tuning the reference cavity. The slope of the discriminator characteristic is determined by the loaded Q of the reference cavity, which is nominally 900. The second directional coupler in the waveguide network monitors transmitter power output,

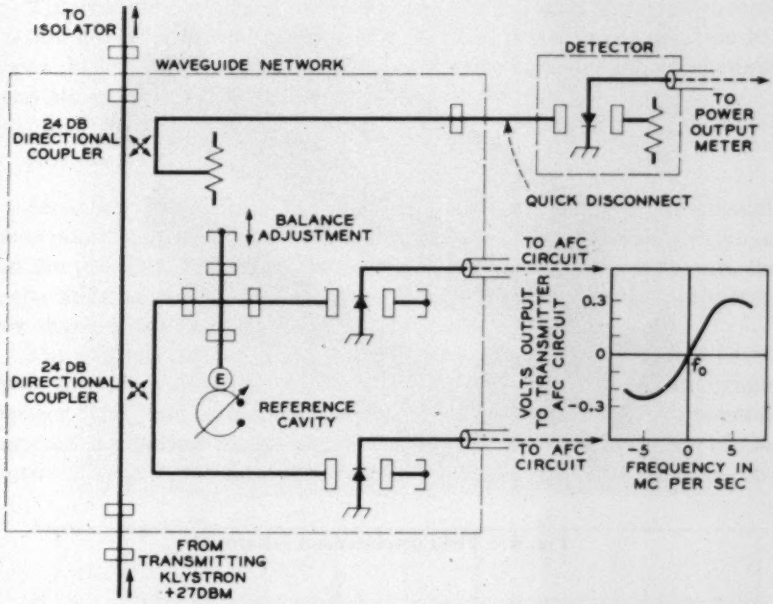


Fig. 7 — Schematic of the transmitter AFC discriminator and power-monitoring network.

and, with a nominal +27 dbm signal from the klystron, the coupler output is +3 dbm. Normally, this signal is fed to a detector and meter circuit that is calibrated in decibels referred to the +27 dbm level. For test purposes, the detector can be rapidly removed, permitting frequency and other transmitter characteristics to be checked.

To minimize the nonlinear effects produced by reflections in the antenna feed, a high-performance ferrite isolator is required between the klystron output and the antenna system. The field displacement isolator[9] shown in Fig. 8 connects to the output port of the AFC waveguide network, and typieal forward and reverse loss characteristics are shown in Fig. 9.

The transmitter output is fed to the channel-combining networks through a waveguide switch, which normally allows the RF energy to pass through unattenuated. When a klystron is replaced, the initial oscillating frequency may be considerably different from nominal. To prevent interference with adjacent channels, the switch can be temporarily closed, thereby attenuating the output signal by more than 80 db.

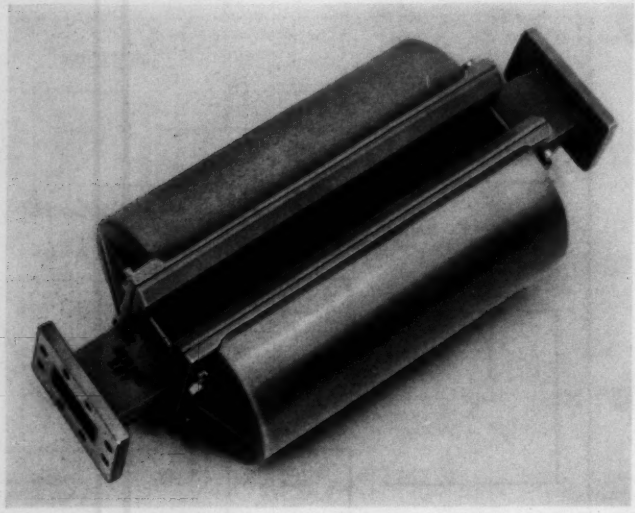


Fig. 8 — Field displacement isolator.

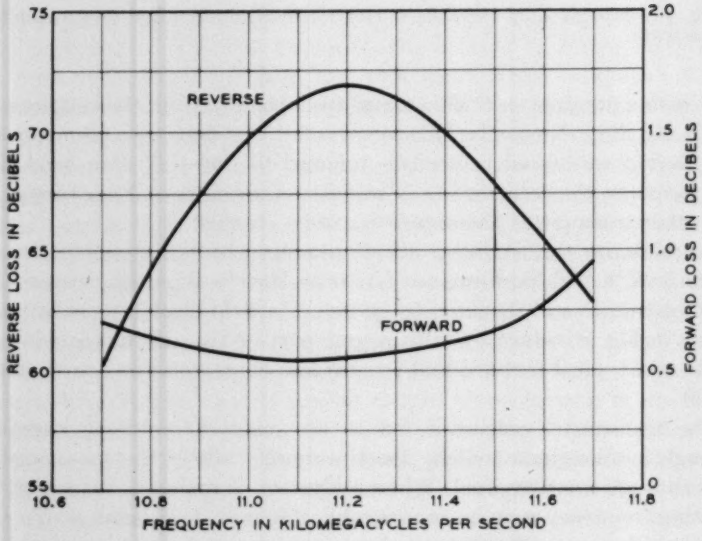


Fig. 9 — Forward and reverse loss characteristics of the TJ isolator.

The channel-combining networks are identical with the channel-separation networks on the receiver (Fig. 4). The principle of operation has been described elsewhere,[10] and representative transmission characteristics are given in Fig. 10.

4.1.2 *Receiver Radio Frequency Units*

Incoming RF channels from the antenna are selected by channel-separation networks identical with the combining networks mentioned in the previous section. In the case of the last receiver in a line-up, the separating network is not required, since at this point the number of RF channels has been reduced to one.

The selected channel is fed to the receiver modulator through a band-pass filter, a waveguide tuner and a critically dimensioned waveguide spacer. The filter serves the two-fold purpose of providing suppression against interfering image signals and enhancing the efficiency of the modulator by reflecting out-of-band modulation products back into the

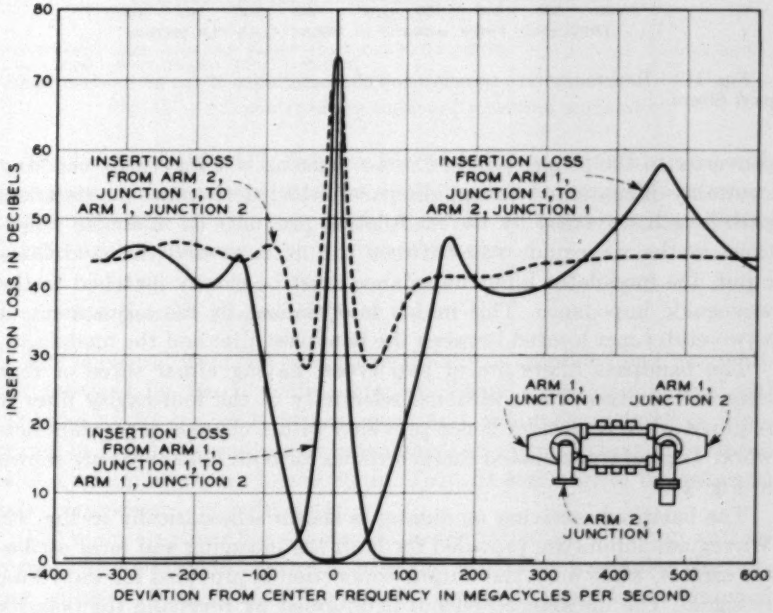


Fig. 10 — Representative transmission characteristics of the RF channel separation-combining networks.

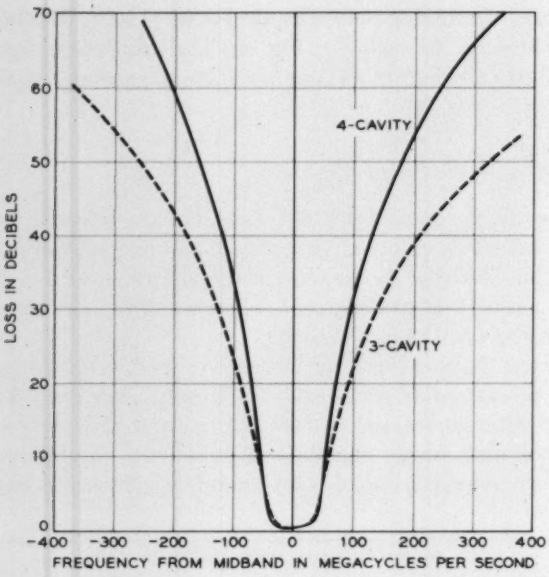


Fig. 11 — Representative transmission characteristics of the RF receiver bandpass filters.

converter in the proper phase. Correct phasing is achieved by choosing a suitably dimensioned waveguide spacer, which determines the electrical path length traversed by the modulation product. To minimize reflections in the waveguide run between the antenna and the modulator input, the modulator input impedance must be closely matched to the waveguide impedance. This match is optimized by the adjustment of a two-stub tuner located between the bandpass filter and the modulator.

The bandpass filters are of two types, having either three or four resonant cavities. The additional selectivity of the four-cavity filter is required when a receiver is not provided with a channel-separation network. Typical transmission characteristics for both filter types are shown in Fig. 11.

The balanced receiving modulator is shown schematically in Fig. 12. Waveguide inputs are provided for both the incoming and local oscillator signals, while a coaxial output connection is provided for the 70-mc IF signal. The unbalanced output is obtained by reversing the polarity of one diode relative to the other in the hybrid junction assembly, thus permitting paralleled unbalanced output connections.

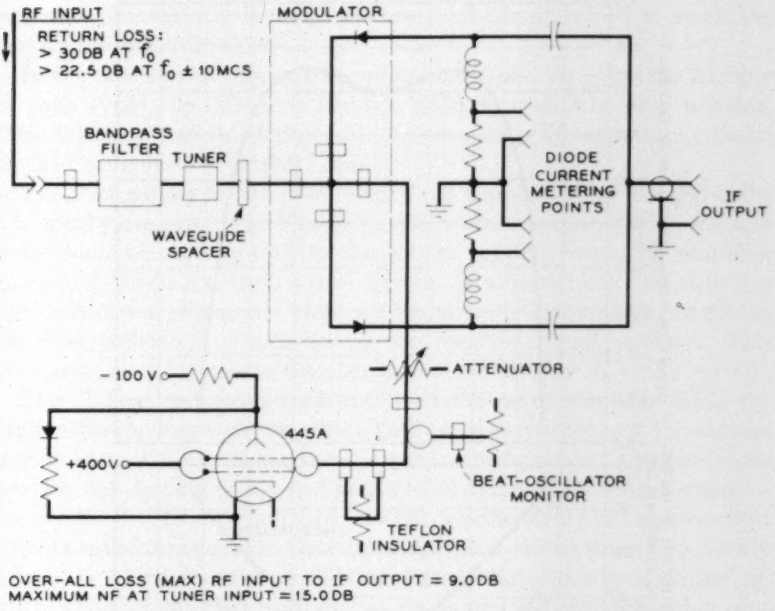


Fig. 12 — Schematic of the balanced receiving modulator.

The local oscillator is a Western Electric 445A klystron identical with that used in the transmitter. It feeds the modulator through a monitoring coupler and an attenuator, which can be adjusted to give a power input to the modulator of approximately 0 dbm.

4.1.3 *Intermediate Frequency Units*

The IF output from the receiver modulator is amplified in a preamplifier followed by the IF main amplifier. These units have a passband centered at 70 mc with typical gain and delay characteristics illustrated in Fig. 13. Input and output impedances are 75 ohms unbalanced with a minimum return loss of 20 db. Table II summarizes the gain and tube complement for both units.

The main amplifier has a nominal power output of +5 dbm, and internal automatic gain control on the first six stages maintains the output within 4 db of nominal for a 40-db change in input level. Matched double-tuned coupling circuits are used in all IF units, with nonadjustable auto transformers in the interstage networks. The only adjustable

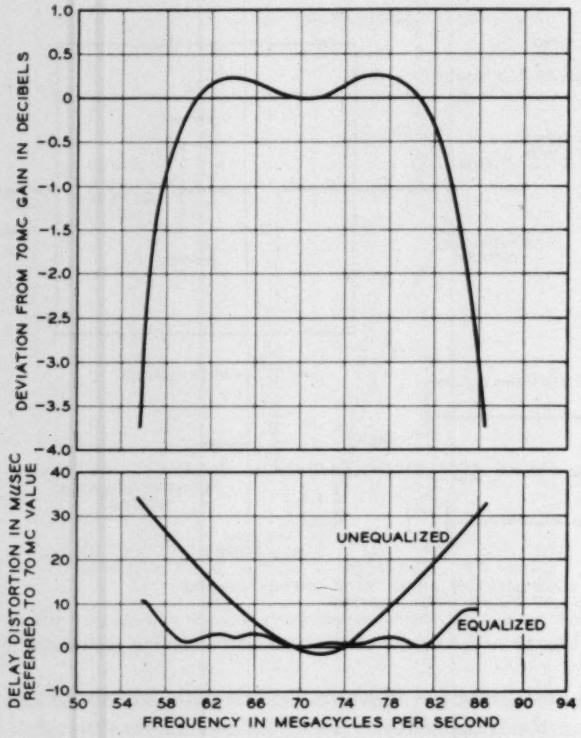


Fig. 13 — IF gain and delay characteristics.

elements are in the input and output coupling networks, which are adjusted for maximum return loss. The marked reduction in the number of tuning adjustments simplifies testing and maintenance, while reducing the possibility of maladjustment. Alternative coupling networks providing higher gains per stage were considered in the initial design

TABLE II

Unit	Gain	Tube Complement
Preamplifier	32 db	two 417A triodes two 435A tetrodes
Main amplifier	75 db	seven 435A tetrodes

but were discarded in favor of the matched circuit, which is much less sensitive to tube changes.

In the interests of minimizing maintenance and reducing the number of tube types, the Western Electric 435A is used wherever possible. This tube is a tetrode of proven integrity with a life expectancy based on field experience of 50,000 hours.

The plate supply for the preamplifier and main amplifier is +200 volts. The total plate current for all tubes in the main IF amplifier is fed through a common resistor located in the power supply. Since the amplifier gain, and hence the total plate current, is automatically adjusted to compensate for changes in received signal level, the voltage across the common resistor is a measure of the received signal strength. This voltage is used to actuate the comparator circuit in a diversity system.

The 75-ohm unbalanced output from the main amplifier feeds the limiter-discriminator circuit. Plate limiting is employed in a three-stage circuit, using 435A tetrodes. Low-forward-impedance gold-bonded diodes provide the clipping action and give a total dynamic limiting of approximately 40 db. The discriminator uses two separately driven antiresonant circuits tuned to approximately 52 and 87 mc. When these circuits are driven by the frequency-modulated signal, amplitude modulation results, which is detected by a double diode and fed through video coupling circuits to the balanced receiving baseband amplifier.

4.1.4 *Baseband Units*

A maximum peak-to-peak signal of 10 volts is required to modulate the repeller of the transmitting klystron, to provide ±4 megacycles deviation. This is provided by a two-stage, balanced video amplifier with an optional input impedance of 75 ohms unbalanced or 124 ohms balanced. The transmitting amplifier utilizes four Western Electric 437A triodes and has a nominal voltage gain of 37 db, adjustable over a range of ±3.5 db. A peaking circuit in the amplifier input provides low-frequency phase correction for a complete repeater when it is used for television transmission. The peaking circuit is provided as a wiring option and is not used in telephone applications.

The receiver baseband amplifier is a single stage balanced circuit utilizing Western Electric 417A triodes. Its principal function is to convert from the relatively high output impedance of the discriminator to the line impedance. This may be 124 ohms balanced or 75 ohms unbalanced in accordance with the wiring option chosen within the unit. With an FM deviation of 8 mc peak-to-peak, the discriminator-balanced

amplifier combination gives a minimum balanced output of −1 dbm (−1 dbv).

4.1.5 *Automatic Frequency Control*

Automatic frequency control is used on both the transmitting klystron and the receiver local oscillator. In the former case the control is electromechanical; in the latter, it is electronic.

The transmitter frequency-error signal is derived from the RF discriminator network described in Section 4.1.1. This error signal is applied to the AFC circuit shown in functional schematic form in Fig. 14. When the frequency error exceeds 300 kc, the meter-type relay operates one of the mercury relays, causing the drive motor to mechanically tune the klystron cavity in a direction to reduce the error. The drive motor then continues to operate until the meter relay is reset. If the frequency error still exceeds 300 kc, the same sequence of events will be repeated until the error is less than 300 kc. The accuracy of the AFC system is controlled by the stability of the reference cavity in the RF discriminator circuit. Over the ambient temperature range from −20 to +120°F the transmitter frequency is maintained within ±0.03 per cent of its nominal value.

Electromechanical control of the transmitter frequency is preferred to electronic control, since most of the frequency error results from changes in ambient temperature. Mechanical tuning corrections for this type of frequency error have much less effect on modulation linearity than does a corresponding correction produced electronically.

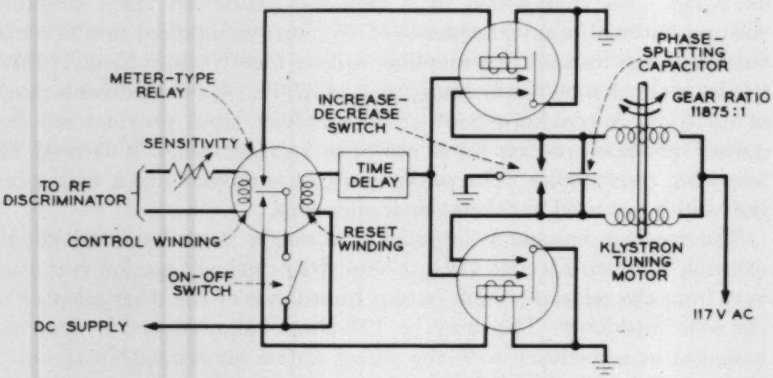


Fig. 14 — Functional schematic of the transmitter AFC circuit.

Automatic frequency control on the receiver is performed electronically since, in this case, the beat oscillator operates at a fixed frequency and modulation linearity is not of importance. The IF error voltage is derived from the signal discriminator, as illustrated in Fig. 15. The error signal is amplified and applied in series with the beat oscillator klystron repeller voltage with the appropriate sign to reduce the error. This automatically maintains the intermediate frequency at the cross-over frequency of the discriminator. If the cross-over point is not exactly at 70 mc, an adjustable bias is provided at the input of the AFC amplifier, which permits adjusting the intermediate frequency to exactly 70 mc. To limit the frequency excursions of the beat oscillator and prevent the receiver's locking onto unwanted signals, a clamp circuit is included in the AFC loop. An additional clamp circuit for the protection of the klystron beat oscillator ensures that the magnitude of the negative repeller voltage never drops below 40 volts.

The frequency of the receiver beat oscillator may be above or below the incoming signal frequency, depending on the channel number. These two conditions correspond to a phase reversal at the discriminator output, and must be compensated for to ensure stability in the AFC feedback loop. The phase correction is obtained by reversing the balanced connections between the limiter-discriminator chassis and the receiver baseband–AFC unit.

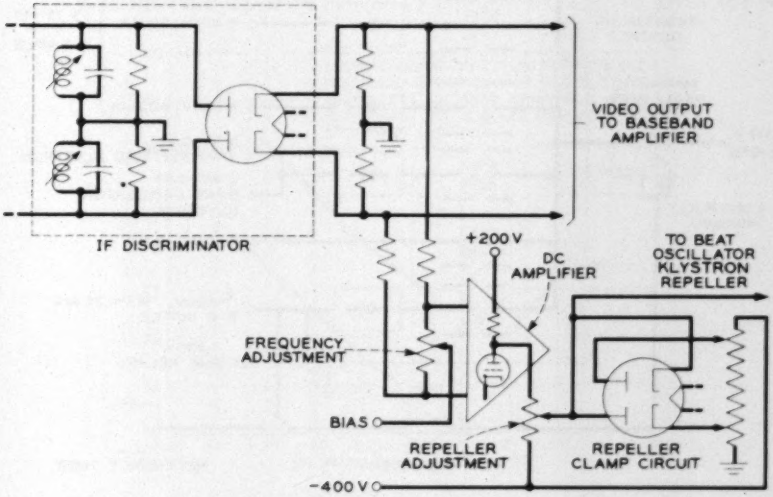


Fig. 15 — Functional schematic of the receiver AFC circuit.

4.1.6 *Power Supplies*[11]

All dc voltages and ac filament supplies are derived from an ac-operated power supply in the lower third of the transmitter-receiver bay.

The principle of operation is illustrated in the block schematic of Fig. 16. Only the −400 volt output is directly controlled by the feedback loop; regulation on the other outputs is dependent on their tracking the −400 volt output. This is achieved by suitable design of the filter networks and the rectifier regulation characteristics. A sensing relay on the −400 volt output ensures that negative repeller voltage is applied to the klystrons before the positive resonator voltages.

Silicon diodes are used throughout as the rectifying elements, so that the only active devices in the circuit are two Western Electric 310A dc amplifier tubes and two voltage reference tubes (423A and 427A) in the feedback control loop. Compared with the more usual series tube regulation, it is expected that the combination of solid state rectifiers and long-life control tubes of proven integrity will greatly reduce power consumption and the annual charges associated with the maintenance of the series tube type of regulated power supply.

The stability of the dc output voltages over the ambient temperature

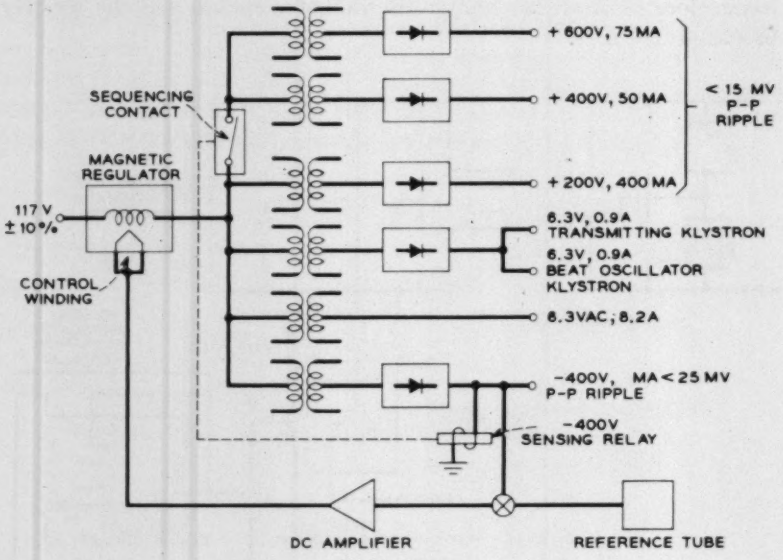


Fig. 16 — Power supply block schematic.

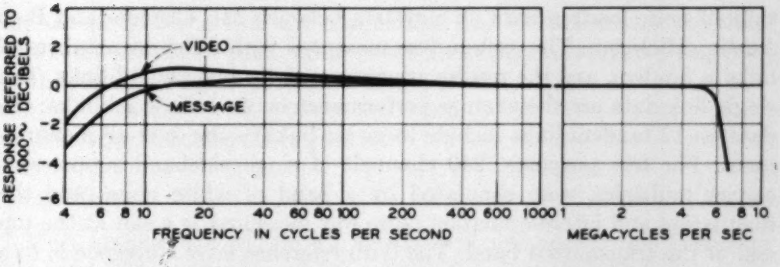


Fig. 17 — Gain-frequency characteristic of a TJ repeater.

range 32° to 120°F and with line voltage variations of ±10 per cent is better than ±0.5 per cent.

4.1.7 *Over-All Performance Characteristics*

The gain-frequency characteristic of a single TJ repeater relative to 1 kc is shown in Fig. 17. The absolute gain of a repeater is nominally 16 db measured from transmitter input to receiver output between 124-ohm balanced impedances. In diversity applications, 6 db of this gain is lost when two transmitters are fed simultaneously resulting in a net gain of 10 db between HP FL IN and HP FL OUT on the diversity switch unit.

To illustrate the linearity of the TJ equipment, Fig. 18 gives the re-

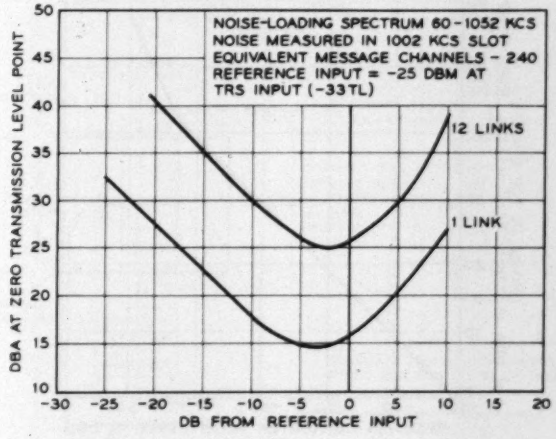


Fig. 18 — Typical noise loading performance of a TJ system.

sults of noise loading tests on a system between Mt. Clemens and Port Austin, Michigan. The system was measured with different numbers of links in tandem, and the results are presented for one and 12 links. The single-link data are the average performance on six individual links; the data on 12 tandem links include these six links in the over-all measurement. For test purposes, 240 channels of single-sideband suppressed-carrier multiplex were simulated by a band of white noise, and the fluctuation and intermodulation noise was measured in a slot at the top end of the transmitted band. The 0 db reference level corresponds to a total input noise power of −25 dbm at HP FL IN on the diversity switch and transmission unit, which is a −33 db transmission level point.

The contribution of fluctuation-type noise to total noise at the receiver output will depend on the received RF carrier level. This relationship is illustrated in Fig. 19, where the noise is given in dba for the top channel in a system carrying 240 channels of suppressed-carrier type-L multiplex. The dba readings are based on a peak-to-peak frequency deviation of 8 mc and a load capacity rating for 240 channels of 21 db.

In television transmission, the linearity characteristics of interest are differential gain and phase. Without pre-emphasis or delay equalization, the differential gain and phase on a single link are respectively 0.5 db and 4.5°. With 13.2 db of pre-emphasis and IF delay equalization, the corresponding figures are 0.1 db and 0.5°.

Low-frequency noise objectives at the output of a TJ receiver require

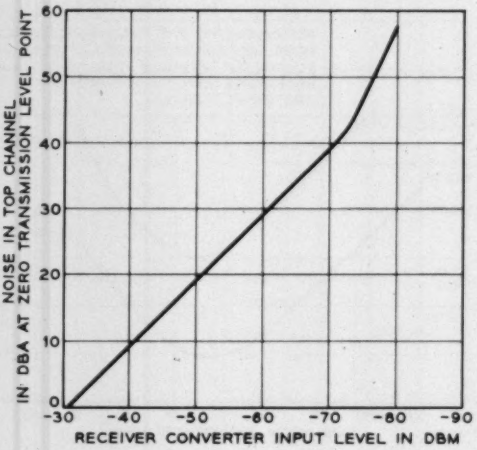


Fig. 19 — Fluctuation noise output as a function of received signal level.

the use of dc heaters on the transmitter and local oscillator klystrons. The ratio of peak-to-peak signal to low-frequency unweighted rms noise at the receiver output on one TJ link is 67 db.

4.2 *Diversity Switching Arrangements*

4.2.1 *General*

Most applications of the TJ system will use one-for-one frequency diversity. This provides protection against multipath fading and equipment failures, and provides alternate facilities during maintenance and equipment additions. At each repeater the diversity equipment selects one baseband output from two receivers and applies this signal simultaneously to two transmitters. The selection of a receiver is controlled by a logic circuit, which, along with the switch, is contained in the diversity switch and transmission unit.

4.2.2 *Diversity Switch and Transmission Unit*

Fig. 20 is a block schematic of the diversity switch and transmission unit as used at a telephone repeater. The audio and high-frequency components of the baseband signal are separated and subsequently recombined by high-pass–low-pass filters. This permits connection to the D-type alarm and order-wire equipment without affecting through trans-

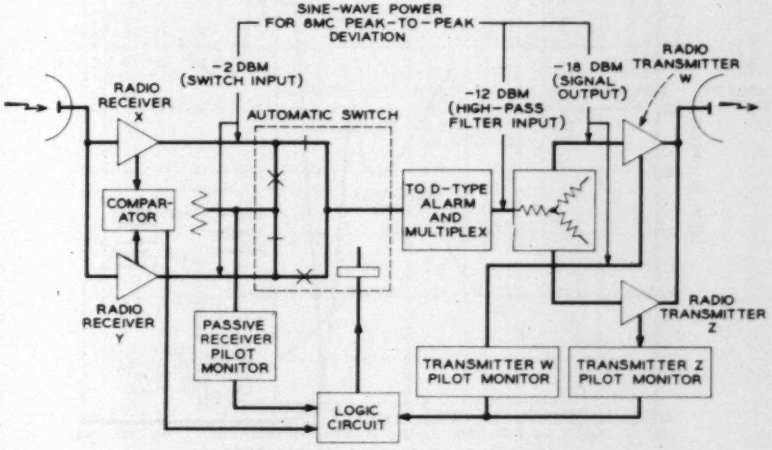


Fig. 20 — Block schematic of the diversity switch and transmission unit

mission of the higher frequency multiplex signals. Furthermore, if the multiplex signals are connected through dropping equipment that does not normally pass low frequencies, the separation filters enable the audio portion of the band to be bypassed without undue attenuation. Typical loss-frequency characteristics for the split-apart filters are given in Fig. 21. Adjustable attenuators and pads provide for equalizing receiver levels and setting frequency deviations on the transmitters.

On a fully developed TJ route the order-wire and alarm information will be carried on only one of the three diversity pairs. In general, therefore, split-apart filters are not required on switch units associated with the second and third systems. In these cases, the switch output can be connected directly to the Y-pad input through a 10-db pad. An exception to this rule occurs with multiplex dropping facilities, as described previously.

The signal switch is controlled by a logic circuit operating under instructions from transmitting and receiving pilot monitors and an RF signal comparator. In general, three monitors are utilized to sense the presence of a 2600-cps pilot tone at the two transmitter AFC discriminators and at the output of the idle receiver. The pilot tone is applied

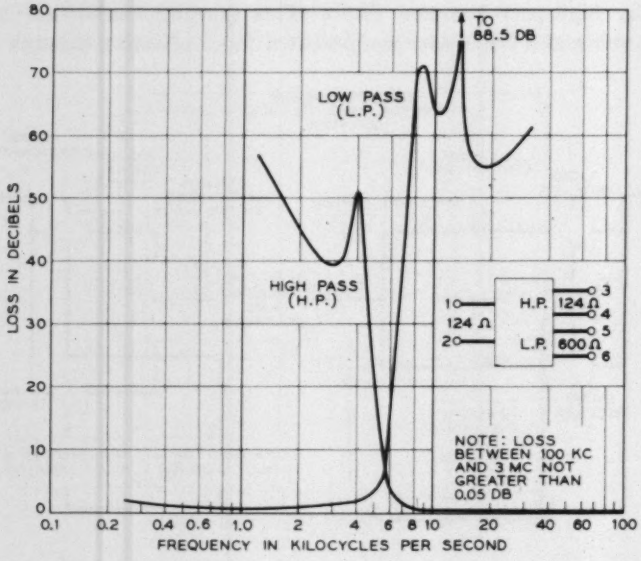


Fig. 21 — Typical loss-frequency characteristics of baseband split-apart filters.

continuously from one end of the system and is looped around at the far end terminal to provide pilot tone on the return path. The 2600-cps tone is also used for signaling purposes, and alarm functions are performed by interrupting the tone temporarily. This arrangement provides a fail-safe alarm system, which will be described in more detail in Section 4.4.

The functional schematic of a pilot monitor and its selectivity characteristic are shown in Fig. 22. During signaling, the pilot tone is applied to the line intermittently, and enough time delay must be built into the monitors so that the absence of tone between pulses may be ignored.

The RF signal comparator is actuated by voltages proportional to the IF plate currents in the X and Y receivers. As described previously, these voltages are an indirect indication of the received RF signal strength. Fig. 23 shows a simplified schematic of the comparator, which provides an output to the logic circuit in the form of high and low contacts on the comparator relay C. Diodes CR1 and CR2 act as clamps on the comparator tube grid voltages, so that the circuit is inactive until one or other of the received RF signals reaches a level of approximately −40 dbm. The operational characteristic of the comparator as a function of input signal

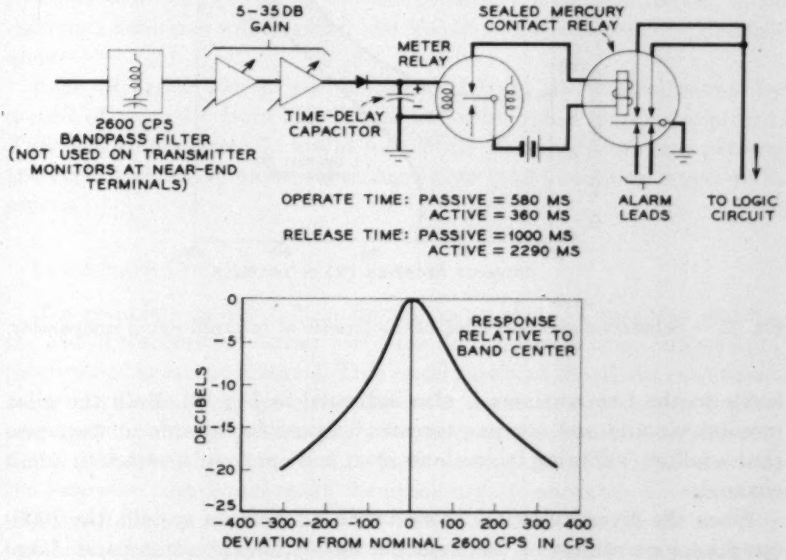


Fig. 22 — Functional schematic and selectivity characteristic of the 2600-cps pilot monitor.

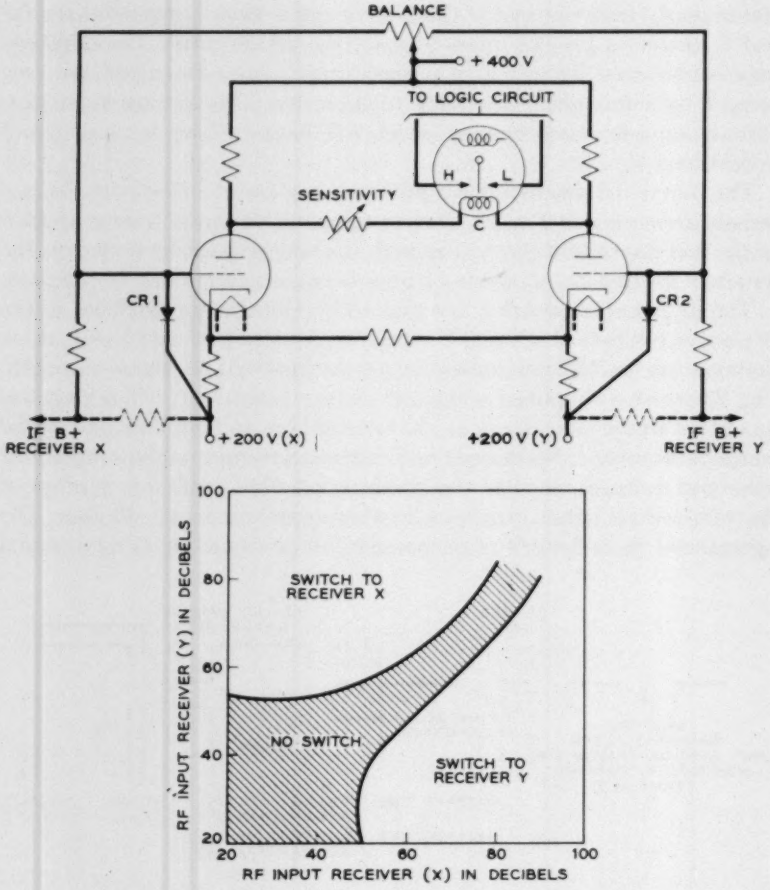


Fig. 23 — Schematic and operating characteristic of received signal comparator.

levels to the two receivers is also indicated in Fig. 23. Both the pilot monitor circuits and comparator are designed to fail-safe in the sense that a failure will bring in an alarm or at least prevent a switch to a bad channel.

When the diversity switch is used with a television system, the 2600-cps pilot tone cannot be used, since it lies within the video band. Likewise, the band-separation filters cannot be used. As a result, the diversity

switch is controlled entirely by the comparator relay, protecting the system against selective fading but not against all equipment failures.

4.2.3 *Diversity Switch and Logic Circuit*

The signal-switching function is performed by a wire-spring relay with make-before-break contacts. During switchover, the relay contacts are momentarily bunched (for approximately one millisecond) paralleling the outputs of the two receivers. If the receiver outputs are equal in magnitude and phase, connecting them in parallel produces no level change, since they are at the same potential. With selective fading the condition of equal baseband levels during a switch is generally met, since the fade must be quite deep before it affects the receiver output. Thus, with proper adjustment, switches due to selective fading are hitless and will not result in data transmission errors. Likewise, manual switches made during system maintenance will also be hitless.

In the case of switches resulting from an equipment failure, the situation is different. Due to the time delay built into the pilot monitors, 2.3 seconds are required to recognize the trouble and make the switch. During this period there will be no transmission and errors in data systems will result. This is not an extremely serious situation, since equipment failures will be much less frequent than atmospheric disturbances.

Each diversity switch can be operated from the control center by means of an order from the D-type order-wire and alarm equipment. This feature is extremely useful for maintenance purposes and permits the rapid location of any level changes that might occur on a system in service.

4.2.4 *Squelch Circuit*

If a complete power failure should occur at a TJ repeater station, the loss of transmitted carrier will cause an increase in noise output from receivers at adjacent stations. This could result in the alarm equipment being unable to interrogate the system to locate the station in trouble and might require a visit to every repeater on the route. Furthermore, if the system connects to a long-haul system carrying other circuits, the excessive noise could make the other circuits unusable. To eliminate this possibility, a squelch circuit is available that cuts off the receiver baseband amplifier when the received RF signal drops below a prescribed value. The circuit is actuated by the IF plate current in the same manner

as the comparator. When the plate current exceeds a certain value, plate voltage is removed from the receiving amplifier tubes. The depth of fade at which the squelch circuit operates is adjustable, but is generally set to actuate at a received RF level of −80 dbm.

4.3 *Nondiversity Applications*

Situations will arise where diversity operation is not required, most commonly in television applications. To facilitate interconnection between transmitters and receivers when the diversity switch and transmission unit is not required, a special connecting panel is available to provide the necessary attenuators, equalizer, cable terminations and access jacks.

4.4 *Order-Wire, Alarm and Control System**

The TJ repeater stations are usually unattended. For this mode of operation to be feasible, attended points responsible for the maintenance of the system must have information promptly as to equipment failures and other abnormal conditions occurring in the unattended radio stations. The type-D alarm, control and order-wire system was developed for this purpose. As the name implies, this system also provides the voice-communication facility between stations and remote control from an attended point of certain functions at the unattended stations.

The principal features of the D2 system may be summarized as follows:

i. It transmits up to 18 distinct alarm functions and six other indications from each unattended station to its associated alarm center. The indications are generally used to determine which receiver of a diversity pair is in use.

ii. It transmits 11 remote control orders from the alarm center to each unattended station under its control.

iii. It allows a maximum of 14 unattended stations to be associated with one alarm center "main station."

iv. It provides order-wire talking and monitoring facilities at each radio station and at the alarm center, with extensions to other points as required. The order-wire channel between radio stations is transmitted in the baseband below 4 kc. All points having access to the circuit are linked together on a multistation or "party line" basis.

v. It provides alarm and control signaling by means of a single frequency tone of 2600 cps within the same four-wire circuit used for the order wire, so that no line facilities other than the voice-frequency chan-

* This section prepared by H. H. Haas.

nel of the radio system are required, except for extensions to points off the radio system.

4.4.1 *Alarms*

The alarm center (main station) signals the remote points (substations) by pulsing the 2600-cps tone. Signal receivers, bridged on the outgoing line at each substation as shown in Fig. 24, transform these tone pulses into dc pulses to operate decoding relays. Each substation is capable of recognizing pulse codes directed to it and translating these codes into orders or, alternatively, answering codes that are essentially queries by reverting certain received pulses.

The main station transmits a continuous tone in the idle condition through each intermediate substation. The normally closed pulse-reverting path of a terminal substation sends the tone back to a signal receiver at the main station. This closed loop is self-alarming in the event of failure of the alarm line and, in addition, provides the means of alerting the main station when a trouble occurs at any substation.

When an alarm condition occurs, substations initiate automatic interrogation from the main station by inserting a 2600-cps stop filter in the return line until this process is completed. A relay sending director circuit at the main station reacts to the interruption of the tone loop by sending out a train of pulses to successively identify the substation at which the alarm occurred and the nature of the trouble — that is, which of the 18 alarms exist at that station. The substation provides automatic identification and scanning by reverting a unique combination of the received pulses, and momentarily closes the pulse-reverting path from outgoing to return line at appropriate times, under control of the relay circuit that counts and decodes the received pulses. The information derived from reverted pulses received at the main station is displayed on lamps. An attendant at the alarm center can also scan the alarms at any time by sending out an order.

4.4.2 *Remote Controls*

Basically, remote control techniques of the D2 system are similar to those involved in reporting alarms. Typical orders to a radio station may be for remote manual control of the emergency engine alternator or for protection switching. Other orders are provided to request a scan of alarms and other indications. In the first example, one-way selective signaling is employed; in the second, revertive pulsing conveys the response to the interrogation.

The sending director sends out the station-selection and order code

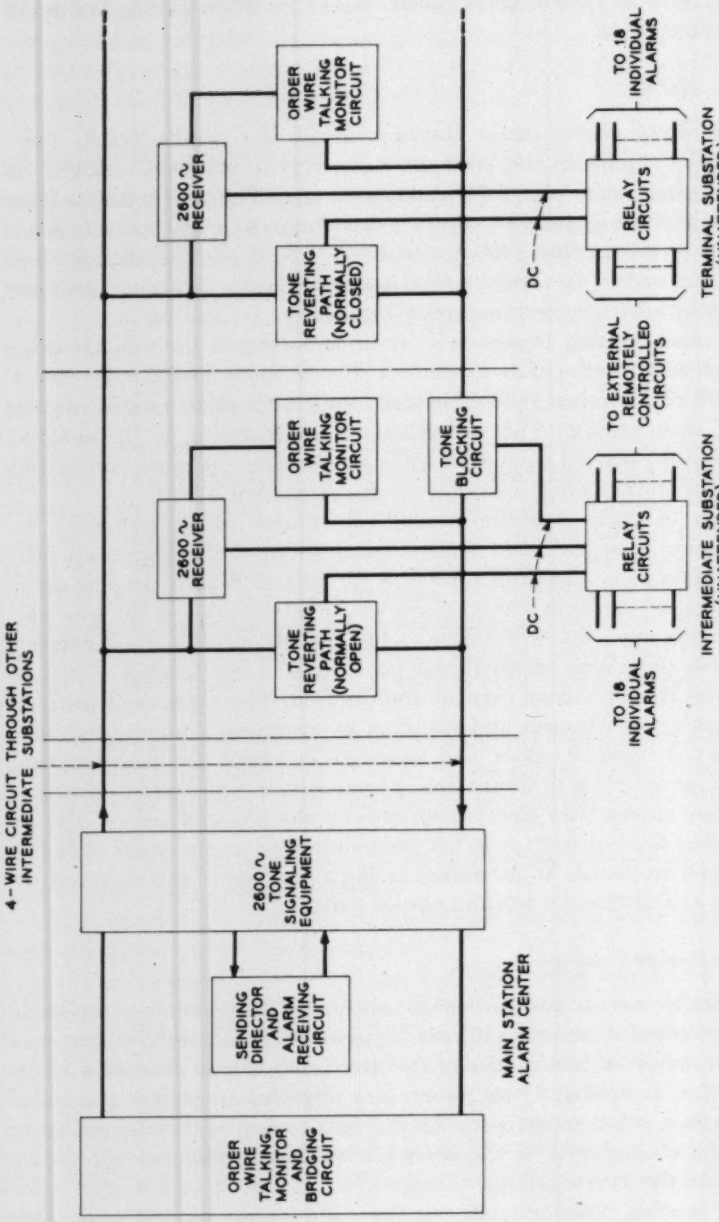


Fig. 24 — Block diagram of D2 alarm and control system.

pulses in response to the operation of keys. To send an order the alarm center, an attendant first operates one of 14 station keys, then one of 11 order keys. The relays at the substations count and decode the pulses. The wanted substation recognizes from the initial pulses that the order applies to it, and proceeds to translate the remaining pulses into one of 11 orders at that particular station.

The 2600-cps tone-signaling equipment of the D2 system is a transistor version of the standard in-band type used on telephone trunks.[12] In the idle "tone-on" state, when no alarm or control pulsing is taking place, the tone level is low and is filtered out of the monitoring circuits of the order wire. The pulsing sequence, although at a higher level, is of such short duration that it can be transmitted over the order wire without objectionable interference with voice communication. Conversely, interference between the voice and signaling equipment is prevented by guard circuits built into the signaling equipment.

Main stations may be located at the focal point of up to four converging radio routes or at an intermediate point along a route as well as at the end of the route as shown in Fig. 24. In addition, a main station can be used to alarm radio routes that converge at unattended substations (junction substations).

4.4.3 *Order Wire*

Way stations on the order wire talk on one side of the four-wire line and monitor on the other. To complete the circuit, the two sides are bridged at the main station so that all intersubstation communication is via the main station.

4.5 *Antenna Systems*

Some of the factors affecting the choice of antennas are; antenna height required, gain, return loss, location and cost. Generally speaking, there are five standard antenna arrangements suitable for use with the TJ system.

4.5.1 *5-Foot Paraboloid*

The antenna shown in Fig. 25, is a rear-feed, ring-focus, dual-polarized paraboloidal antenna. It comprises a 5-foot-diameter paraboloid, a primary feed system, a radome and a ball-swivel base with adjustable support rods. It is primarily intended for use in "periscope" antenna systems, but may be used as a direct radiator.

The feed is made of a straight section of circular waveguide, which

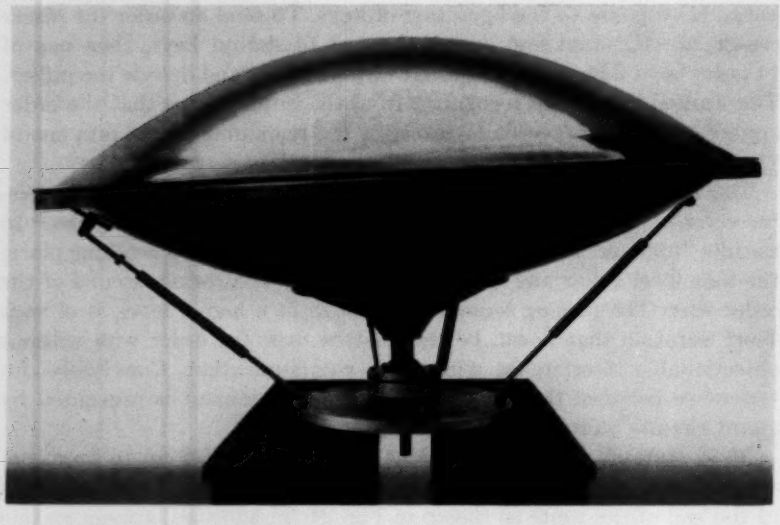


Fig. 25 — TJ 5-foot dual-polarized paraboloidal antenna.

passes through the apex of the paraboloid and along its axis. A specially shaped 3-inch disc reflector is located at the waveguide aperture and serves to direct the energy back on the surface of the paraboloid.

The gain of this antenna is 42.1 db over an isotropic radiator at a frequency of 11,200 mc and varies less than 1 db over the 10,700- to 11,700-mc common carrier band, as shown in Fig. 26. The return loss and cross-polarization discrimination are in excess of 20 db. Fig. 27 shows

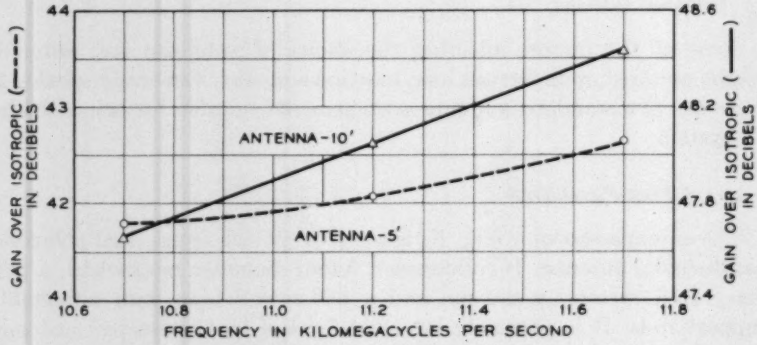


Fig. 26 — Gain-frequency characteristic of the TJ radio 5- and 10-foot antennas.

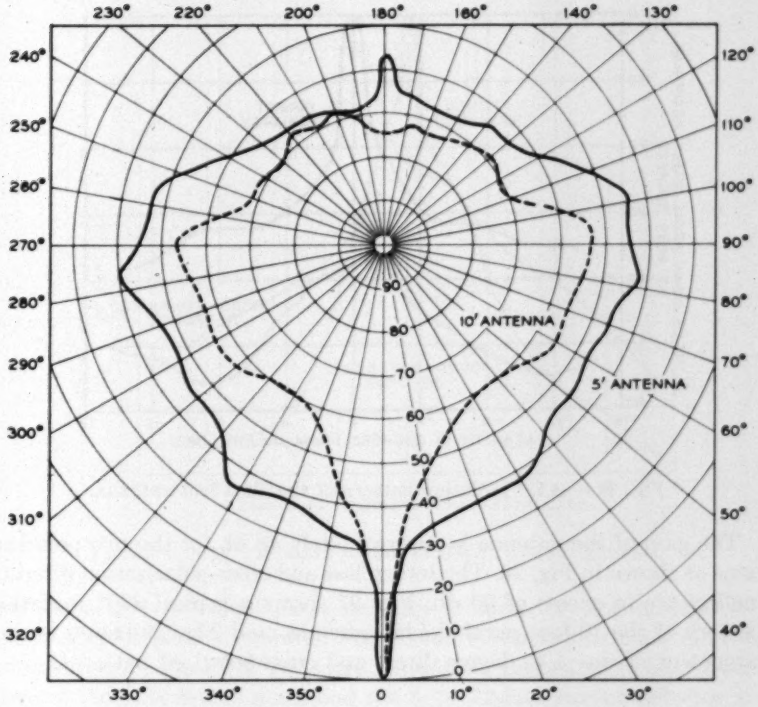


Fig. 27 — Typical 360° radiation pattern for the TJ 5- and 10-foot antennas.

a typical 360° radiation pattern of the 5-foot paraboloidal antenna, and Fig. 28 shows an enlargement of the ±15 degree direct and cross-polarized patterns.

4.5.2 *10-Foot Paraboloid*

Natural elevations are often available in mountainous country for repeater sites. In these cases, path clearance is not usually a problem and the antennas are mounted on short towers whose height is just sufficient to provide foreground clearance. The antenna is composed of a 10-foot paraboloidal reflector, a primary feed system and an A-type mounting frame that provides azimuth and elevation angle aiming adjustments. The feed arrangement is similar to the 5-foot paraboloidal antenna. The 10-foot paraboloid does not have a radome, but heaters are available for both the feed and the paraboloidal reflecting surface.

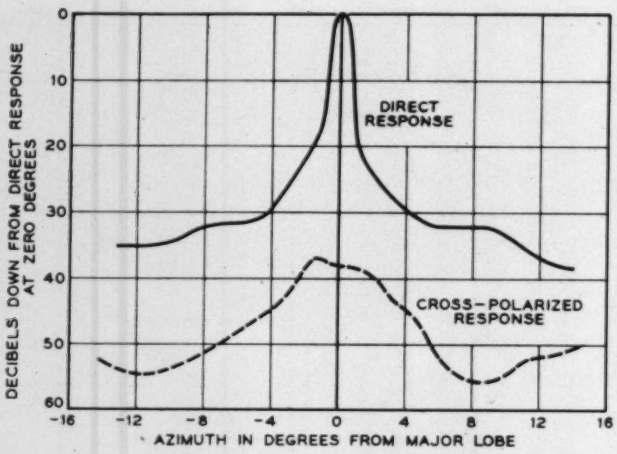


Fig. 28 — ±15° radiation pattern of the TJ 5-foot antenna.

The gain of the antenna is approximately 48 db for the two polarizations as shown in Fig. 26. The return loss and cross-polarization discrimination are in excess of 20 db. Fig. 27 shows a typical 360° radiation pattern of the 10-foot paraboloidal antenna, and Fig. 29 shows an enlargement of the ±15 degree direct and cross-polarized patterns.

4.5.3 *Periscope Antennas*

In many areas of the country the terrain is relatively flat and paths 25 to 30 miles in length require antenna heights in the order of 250 feet. The use of paraboloidal antennas as direct radiators is not desirable, because of the long rectangular waveguide runs and their associated high loss. The horn-reflector antenna, with its 3-inch circular waveguide and combining network performs well at 11,000 mc, but it is too expensive The "periscope" antenna system, which satisfactorily fills this need, consists of a paraboloidal antenna mounted at or near ground level, usually on the roof of the repeater station, illuminating a reflector at the top of the tower. It has the advantages of requiring a minimum length of waveguide and providing an over-all antenna system gain that is equal to or greater than the gain of the paraboloidal antenna alone.

Some of the factors determining the gain[13] of a "periscope" antenna are the frequency, the relative size of the apertures of the paraboloid and the reflector, and the separation between them. Standard arrangements for the TJ system consist of the 5-foot paraboloidal antenna

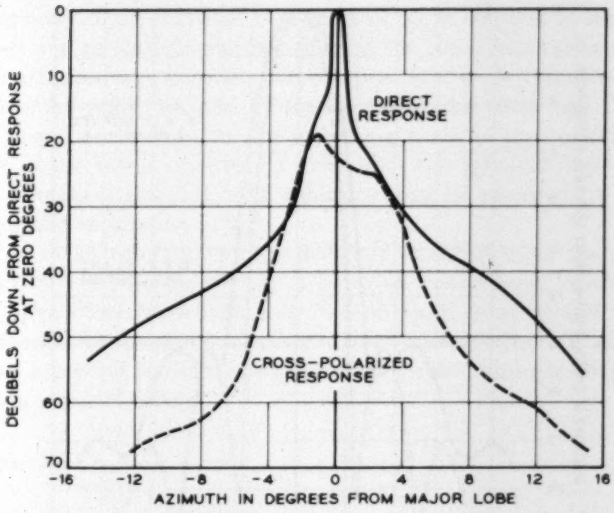


Fig. 29 — ±15° radiation patten of the TJ 10-foot antenna.

and either a 6- × 8-foot plane reflector or an 8- × 12-foot reflector, which may be plane or curved. Fig. 30 shows the approximate gains to be expected for these TJ "periscope" antennas. Estimated radiation patterns of the 6- × 8-foot plane and the 8- × 12-foot curved reflector are shown in Fig. 31.

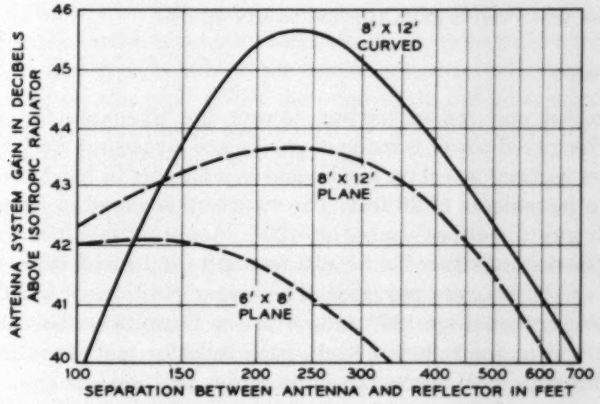


Fig. 30 — Gain of periscopic antenna system.

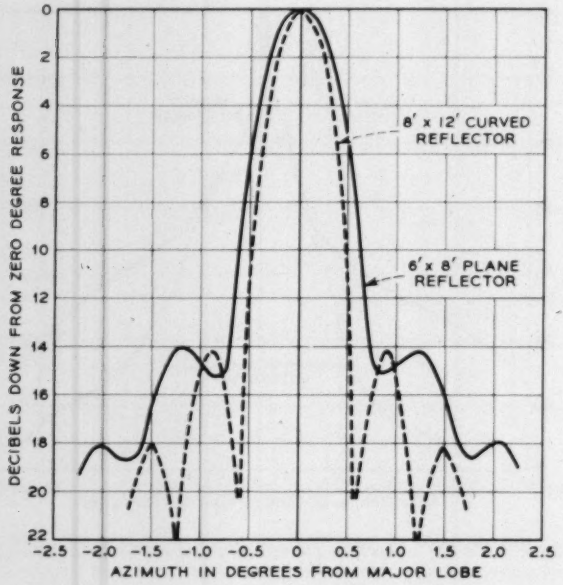


Fig. 31 — Estimated radiation pattern of 6- × 8-foot plane and 8- × 12-foot curved reflectors.

The gains of a "periscope" system for spacings between 25 and 100 feet are not shown in Fig. 30. It suffices to say that the combination of the 5-foot paraboloid and the 6- × 8-foot plane reflector with these separations will yield a gain approximately 42 db.

4.5.4 *Towers*

Both guyed and self-supporting towers are available for the TJ system. The guyed tower is made of galvanized structural steel, triangular in cross section, 4 feet on a side, and is available in heights of 80 to 300 feet in increments of 20 feet. The structure is guyed in three directions with guy directions spaced at 120°. Mounting facilities for three passive reflectors of either the 6- × 8-foot or 8- × 12-foot type, or combinations of the two, are provided. The center positions of the three reflectors are separated by 120° and, with the flexibility inherent in the mounting and in the reflector itself, each reflector may be adjusted in azimuth through ±40 degrees from its center position.

The self-supporting tower is made of galvanized structural steel, square in cross section, with a straight top section and below this a uniformly

tapering body. It is available in heights of 40 to 300 feet in increments of 20 feet and provides mounting rings at the top of the tower to which may be attached any combination of up to four 6- × 8-foot reflectors, 8- × 12-foot reflectors and 10-foot paraboloidal antennas. The antennas may be attached to the mounting rings in any position irrespective of the tower orientation and, with the adjustability inherent in the antennas, two adjacent antennas may be oriented to a minimum angular separation of 25°.

Both of these towers have been designed to tilt no more than $\pm\frac{1}{4}$ degree and to twist no more than $\pm\frac{1}{2}$ degree under wind loading of 20 lbs per square foot (approximately 70 mph), and to withstand winds of 100 mph with the maximum number of antennas attached. Ground wires and rods to protect foundations and anchors from lightning are supplied with both types of towers.

4.6 *Connecting Circuits*

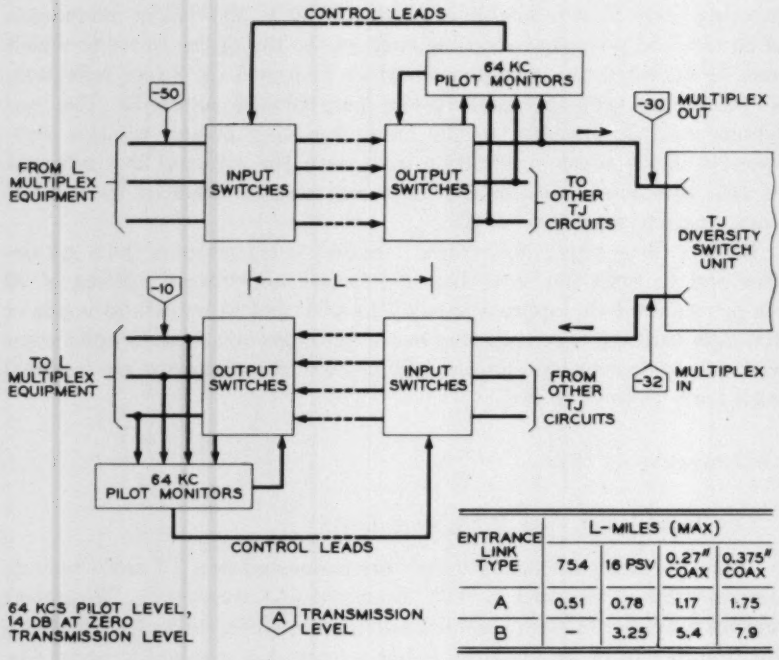
4.6.1 *Type A and Type B Entrance Links*

When L-carrier multiplex signals are connected to a TJ radio system, amplification is required in both directions of transmission. To protect against amplifier failures and facilitate maintenance, protection or standby circuits can be provided on either a manual or automatic basis. The complete ensemble of amplifiers, switches and control circuits is referred to as an *entrance link*. Fig. 32 is a block schematic showing the basic features of an L-carrier–TJ radio interconnection. When fully equipped, three-wire lines with a common protection circuit will supply up to three TJ radio channels. Essentially, the only difference between the type A and type B links is the maximum permissible circuit length. Each circuit on the type A link uses one 40-db flat-gain amplifier and, in the type B link, this is augmented by a second amplifier having a gain shape complementing the cable loss.

4.6.2 *ON-2 Multiplex*

ON-2 multiplex equipment basically provides 48 channels in the 36- to 268-kc band, with a transmitted carrier between each pair of voice channels. For radio applications, modulating equipment is available to stack a second group of 48 channels on top of the basic group to provide a 96-channel system. The frequency band of the upper group is 316 to 548 kc.

The levels out of the multiplex equipment are adequate to drive the



ENTRANCE LINK TYPE	L-MILES (MAX)			
	754	16 PSV	0.27" COAX	0.375" COAX
A	0.51	0.78	1.17	1.75
B	–	3.25	5.4	7.9

Fig. 32 — L-carrier to TJ radio entrance link.

TJ equipment directly, so that no special entrance link facilities are needed. In Table III the channel levels at the point of connection with the TJ equipment are summarized and expressed in terms of the transmitted carrier level at the point in question.

4.6.3 *Television Terminals*

When a television signal is transmitted over radio, it will generally be supplied from a customer or television operating center (TOC) at a

TABLE III

Location	Test Point	
	MX IN	MX OUT
Terminal	−40 dbm	−15 dbm
Dropping point	−34 dbm	−35 dbm

level of 1 volt peak-to-peak (0 dbv). Likewise, at the receiving end of the radio system it must be presented to the customer or TOC at the same level. In most instances, the television terminal equipment and the radio equipment will be physically separated, so that connecting facilities must allow for cable loss and cable equalization. Fig. 33 shows typical connecting circuits when the TJ system is used as an intermediate link between two television operating centers.

Before being applied to the radio equipment, the low-frequency components in the TV signal are reduced in level relative to the high-frequency components by a pre-emphasis network. At the receiving end, a complementary de-emphasis network restores the high- and low-frequency components to their original relative magnitudes. The purpose of the pre-emphasis network is to reduce the modulation index at low frequencies to reduce differential phase and gain distortion.

Alternative connecting circuit arrangements will be used when the television signal is transmitted or received over unbalanced cables. Functionally, they are the same as the balanced arrangements and differ only in the types of networks used for equalization and pre-emphasis.

4.7 *Stand-by Power*

A 5-kw gasoline engine alternator and a 10-kw diesel engine alternator with automatic controls have been made available for use with the

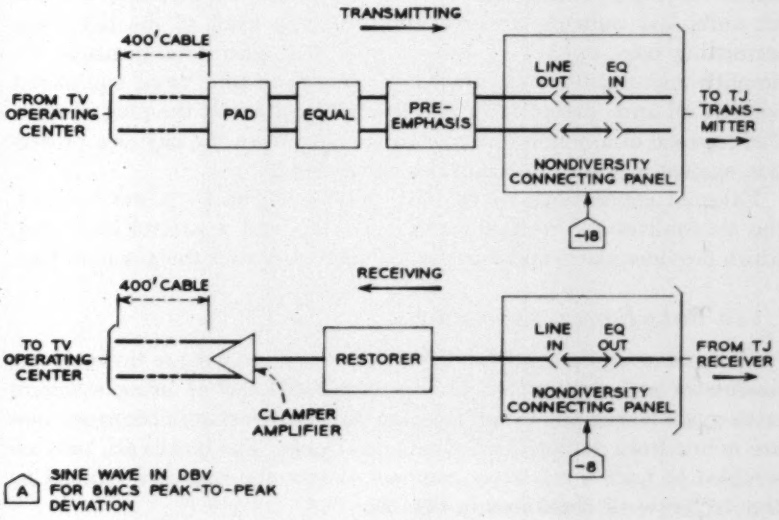


Fig. 33 — Video terminal connecting circuit.

TJ system. When commercial power fails, these units take approximately 30 seconds to reach operating speed and voltage. A further 20 seconds is required for the radio equipment to resume transmission.

For those applications where a 50-second interruption in transmission cannot be tolerated, other types of reserve power plants have been used, which provide essentially no-break operation. This is accomplished by the use of a motor-generator-flywheel combination coupled to a gasoline engine through a magnetic clutch. When commercial power fails, the clutch engages and the flywheel supplies sufficient energy to carry the load and start the gasoline engine.

V. EQUIPMENT FEATURES

5.1 *Transmitter-Receiver Bay*

5.1.1 *General*

The TJ radio transmitter-receiver bay is illustrated in Fig. 34. It consists of a 6-foot floor-supported duct-type framework built to accept standard 19-inch panels. The backplate support for the RF channelizing networks extends above the bay framework, giving an over-all height of 6 feet 8 inches and a total width of $20\frac{1}{2}$ inches.

The upper third of the bay houses the transmitter, while the middle and lower third contain the receiver and power supply, respectively. All units and controls are accessible from the front of the bay, thus permitting back-to-back or back-to-wall floor plan arrangements. To simplify maintenance and reduce out-of-service time from equipment failures, all units except the receiver modulator are of the plug-in type. The removal of any unit automatically shuts down the bay as a protection against exposure to hazardous voltages.

External connections to the bay consist of the ac power conduit, the transmitter and receiver waveguide runs and a control lead plug, which provides alarm and comparator information for the auxiliary bay.

5.1.2 *Radio Frequency Components*

From an RF equipment standpoint, of special interest are the receiving modulator and transmitter AFC networks. Instead of using standard cross-guide fabrication techniques, the waveguide network configurations are milled from sectionalized aluminum blocks. The blocks are then assembled to form a relatively compact, composite structure typified by the AFC network illustrated in Fig. 35.

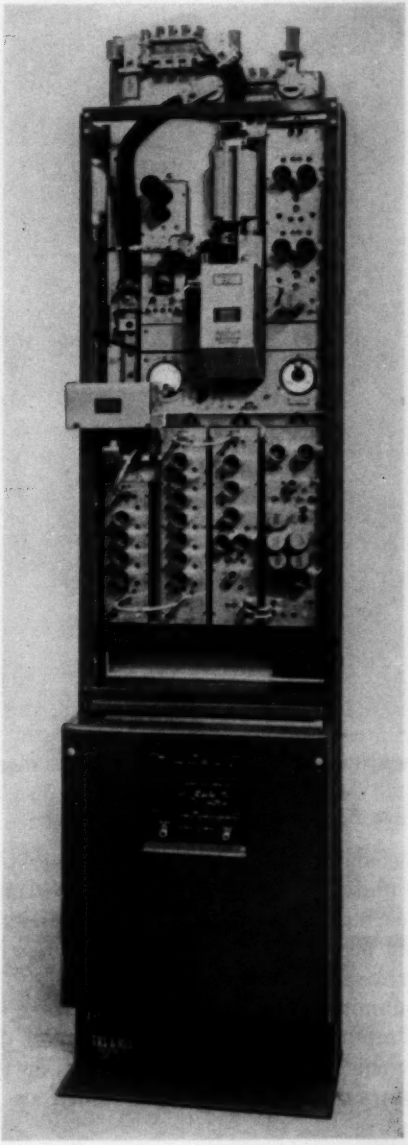


Fig. 34 — TJ transmitter-receiver bay.

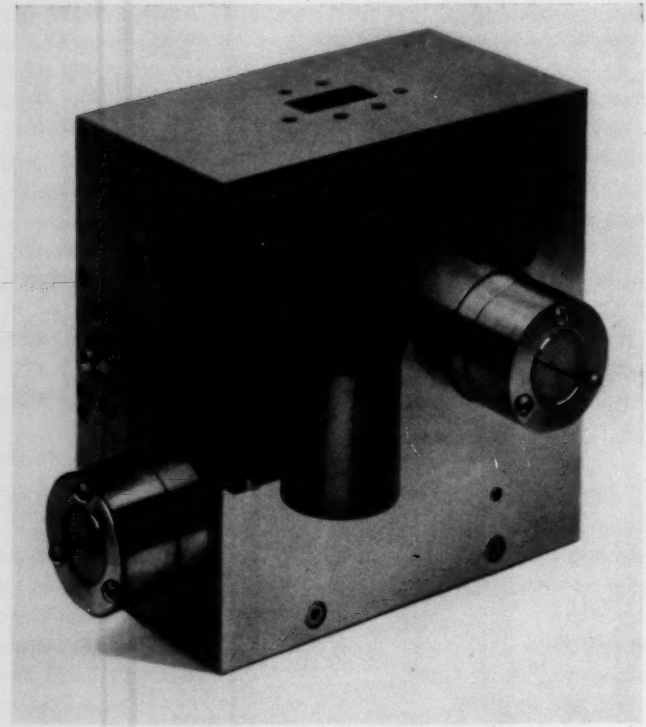


Fig. 35 — Transmitter AFC discriminator and power monitoring network.

5.1.3 *Intermediate Frequency and Baseband Units*

Each IF or baseband unit is assembled in a die-cast aluminum chassis. Typical of the IF units is the main IF amplifier illustrated in Fig. 36. The components for each stage are mounted on an individual plate assembly, which is attached perpendicularly to the tube base of that stage. The individual plates are assembled and wired separately, and then mounted on the die-cast aluminum chassis prior to interstage wiring. In addition to mounting and accurately positioning components, the plate assemblies act as interstage shields. The only adjustments on the IF units are on the input and output coupling networks, which are adjusted for optimum return loss.

The baseband amplifiers use standard assembly techniques, with special precautions being taken to accurately position components. This

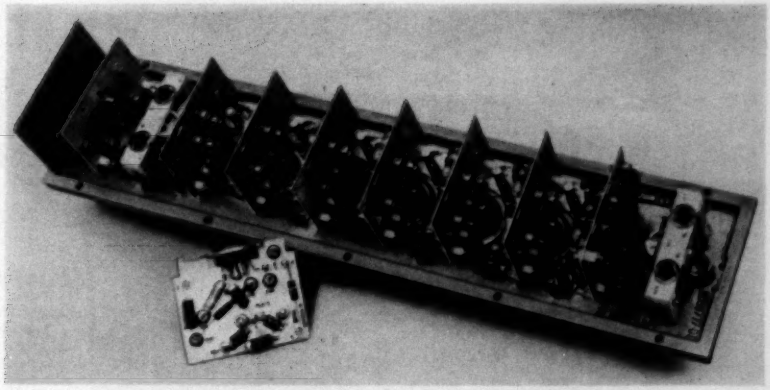


Fig. 36 — Photograph of main IF amplifier.

minimizes variations in parasitic capacities to obtain reproducible high frequency gain characteristics.

All units are equipped with test points for measuring plate and heater voltages. Test points are also provided to measure tube biases for cathode activity tests.

5.1.4 *Power Supply*

An internal view of the TJ power supply is shown in Fig. 37. The hinged front cover provides access to all components and an interlock switch protects against exposure to hazardous voltages. Components mounted in the rear of the unit, can be reached by lowering the backplate, which is hinged near the lower end.

When the radio bay is equipped with a transmitter only or receiver only, the appropriate outputs on the power supply are equipped with bleeder resistors to maintain a constant load on the unit. These loads are located within the power supply and consist of resistors mounted in fuse-clip holders. The power supply also contains all voltage and current metering resistors.

Connections between the power supply and the rest of the radio bay are made through plug-in connectors in the rear of the unit.

5.2 *Auxiliary Bay*

The auxiliary order-wire, alarm and control bay is a 7-foot duct-type framework built to accept standard 23-inch panels. The over-all

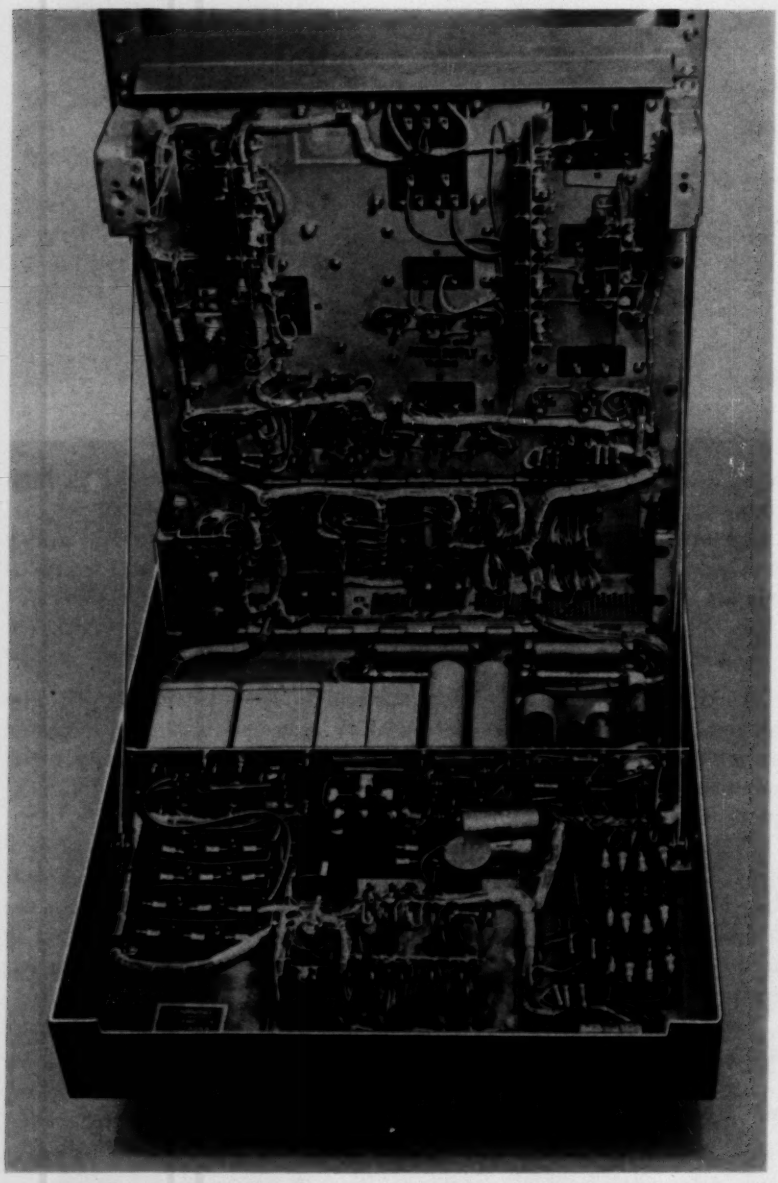


Fig. 37 — Internal view of TJ radio power supply.

height, width and depth dimensions are, respectively, 7 feet, 26½ inches and 16¼ inches.

The bay, which is illustrated in Fig. 38, contains a maximum of six diversity switch units, the D-type alarm equipment and associated power supplies. Like the radio bay, it can be maintained from the front of the bay.

Control lead connections from the radio bays and all external alarm leads are brought to a terminal strip at the top of the auxiliary bay. Connections from the diversity switches are brought to the same panel, permitting cross connections to be made as required in particular applications.

The diversity switch units are built in the form of drawers which can be pulled forward on slides for maintenance purposes. Power and signal connections are made through a flexible cable detail which plugs into sockets in the bay cabling. Switch units can, therefore, be provided or added as needed.

Power rectifiers providing −48 and +130 volts are located below the switch drawers. If these voltages are available from the office supplies, the rectifiers can be omitted and replaced by an optional power-connecting panel.

5.3 *Interconnecting Arrangements.*

5.3.1 *Antenna Feed*

The dual-polarized signal received by a TJ antenna is separated into its two components by a polarizer[14] network, which is illustrated in Fig. 39. The network is located directly behind the antenna and is connected to the radio equipment through two rectangular waveguides. The polarizer network has a cross-polarization discrimination over the TJ band in excess of 40 db, although only 20 db of isolation is realized in practice, due to coupling between polarizations within the antenna.

In a diversity system, the working and protection RF channels are connected to opposite polarizations. This permits maintenance and system additions to be carried out while service is maintained on the opposite polarization. If only one transmitter and receiver are connected to the antenna, the normal channel-dropping network can be omitted by connecting the receiver to one polarization and the transmitter to the other.

The rectangular waveguide connections from the polarizer are connected to the radio bays through a pressure window. Air line connections

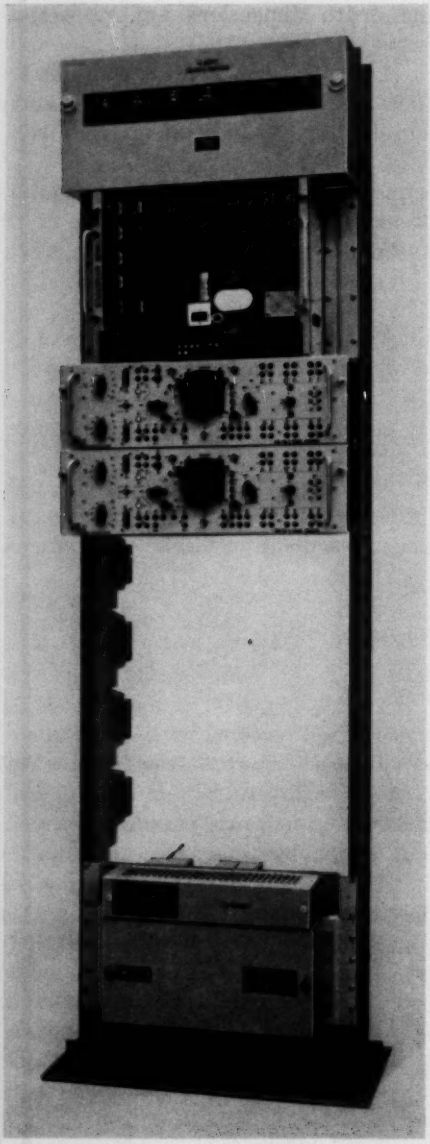


Fig. 38 — TJ radio auxiliary bay.

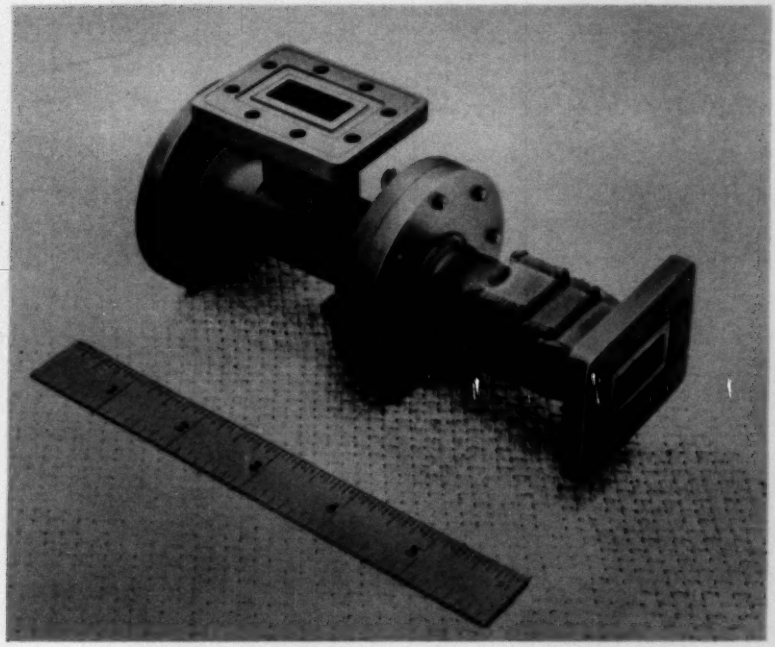


Fig. 39 — Polarization separation network.

to the pressure window from a dehydrator keep the entire waveguide run filled with dry air at a pressure of five inches of water.

5.3.2 *Radio Bay Interconnections*

A typical arrangement for connecting the radio equipment to the antenna system is illustrated in Fig. 40 for a fully equipped diversity system.

When the system has expanded to four two-way radio channels, an isolator is provided between transmitters and receivers, as illustrated for the six-channel case in Fig. 40. The purpose of the isolator is to prevent beat oscillator leakage from a receiver on one polarization going back through the polarizer network and being picked up by a receiver on the opposite polarization. With fewer than four channels the receivers are protected from this type of interference by adequate frequency separation.

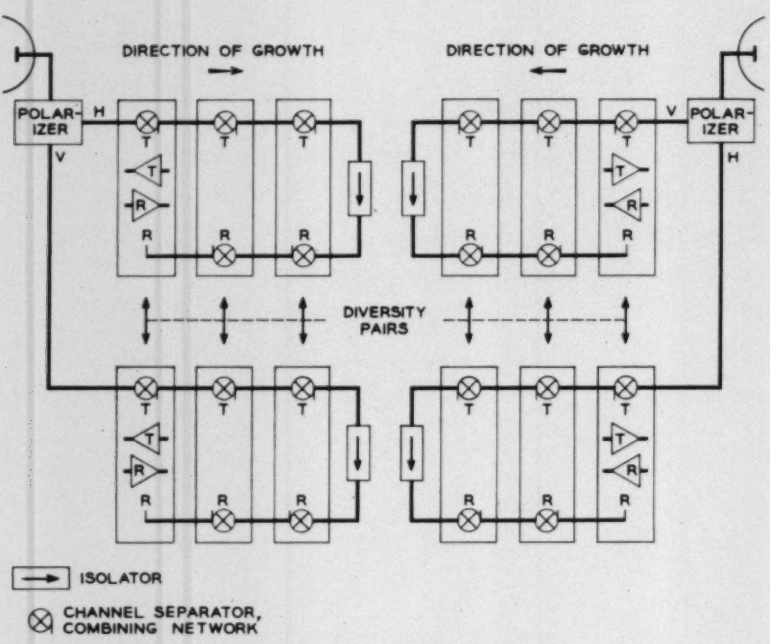


Fig. 40 — Waveguide interconnecting arrangements for a three-channel diversity system.

VI. SYSTEM MAINTENANCE AND TEST EQUIPMENT

Wherever possible, test procedures have been based on the use of commercially available test equipment. The only specialized test sets are:

(a) an impedance matching test set to enable commercial oscillators and voltmeters to be used with TJ circuit impedances;

(b) an IF test set for measuring intermediate frequencies and power output;

(c) an RF test set for generating and measuring 11,000-mc microwave signals;

(d) a transmitter disconnect unit, used to remove one transmitter of a diversity pair from service without introducing sudden level changes in the other which might affect data circuits.

The IF test set is illustrated in Fig. 41. It contains an IF frequency meter and attenuator, together with a calibrated detector for measuring IF signal level.

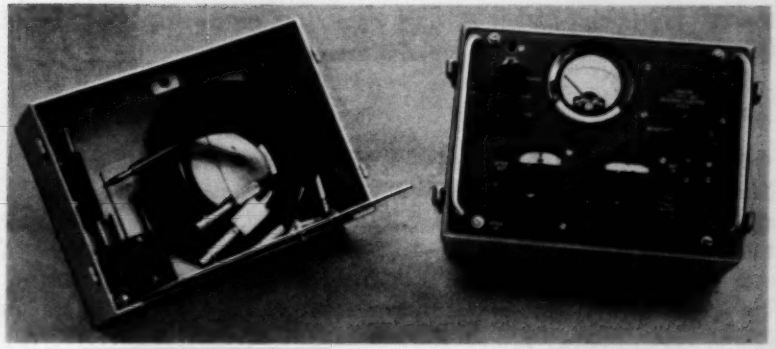


Fig. 41 — IF test set.

VII. APPLICATIONS OF TJ RADIO SYSTEMS

7.1 *Path Selection*

7.1.1 *General*

A number of factors must be carefully considered in engineering any microwave relay system. Before getting down to actual site selection it is necessary to consider all information known about propagation conditions in the general area. For instance, experience has shown that the inverse bending ($k<1$) type of fading is more prevalent in the moist coastal regions, and that some additional tower height should be considered for improved reliability in these areas (k is the radius of a fictitious sphere relative to that of the earth and, under normal propagation conditions, is equal to 4/3). This type of fading is not frequency-selective, and propagation at 4,000, 6,000 and 11,000 mc will be affected similarly.

On the other hand, multipath fading at 11,000 mc is dissimilar to that at 4,000 mc in certain aspects. Multipath fading is due to reflections from stratifications set up in the atmosphere that provide two or more paths of different electrical length for the radio signal. While the time of occurrence and duration of this type of fading at 11,000 mc is approximately the same as it is at 4,000 mc, the received signal levels vary more rapidly at 11,000 mc. Multipath activity is more prevalent at night than during the day; it is greater in warm and humid seasons than in cool and dry seasons; it is greater over water or smooth terrain than

over rough ground and is more severe on longer paths. This type of fading is frequency-selective.

7.1.2 *Path Clearances*

Recommended path clearances for the TJ system are shown in Fig. 42. These clearances are for light-route systems in general and should be applied over true earth. For over-water paths it is recommended that they be increased by $D^2/24$ feet, where D is the path length in miles, in order to provide adequate mid-path clearance for protection against inverse bending type of fades. This suggested increase is based on values of k of approximately 2/3.

7.1.3 *Received Signal Objective*

Based on a 40-db fading margin and a suitable FM breaking allowance, the received signal level objective is −35 dbm to −40 dbm, depending upon the message load. A received signal level of −35 dbm can normally be met for path lengths in the order of 25 miles. Practical

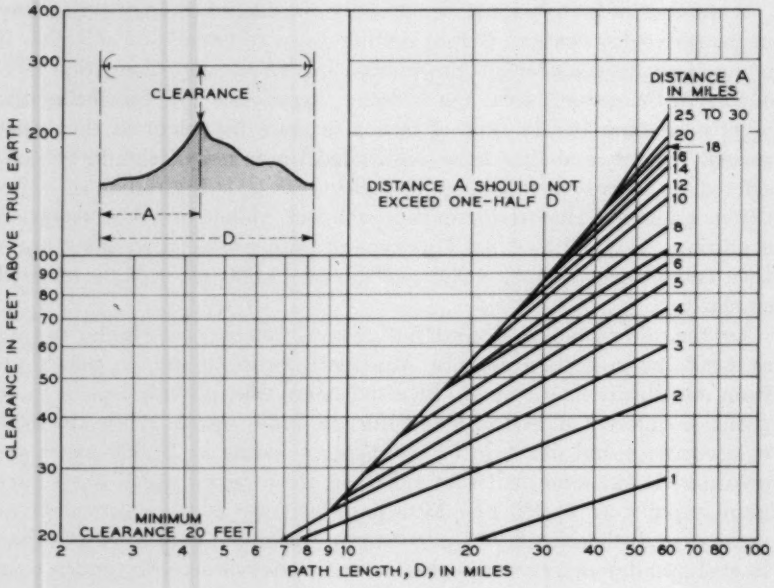


Fig. 42 — TJ radio path clearance requirements.

path lengths will vary from about 10 miles in regions of very severe rain to 40 miles or more in the drier climates.

7.1.4 *Overreach*

Since alternate repeaters use the same "nominal" carrier frequencies for a given direction of transmission, there will be the possibility of a receiver three hops away being subject to interference. The carrier frequencies may differ by as much as a megacycle or more, and the effect of the interference will be to phase-modulate the wanted carrier to produce a tone in the baseband at a frequency equal to the "nominal" frequency difference. Normally, this type of interference is avoided by zig-zagging. In laying out microwave routes, path bearings should be selected so that a total of 50 db discrimination against overreach interference will be provided by the two antennas involved.

7.1.5 *Branching*

The TJ system is designed for a maximum capability of 12 two-way radio frequency channels. For a through repeater and no branching radio routes, six transmitters will normally be used in each direction to provide three working and three protection channels.

Because no frequency is reused at a given repeater, any branching route configuration desired may be used, provided the total number of two-way radio channels does not exceed the system capability of 12. Fig. 43 shows a typical branching or spur route.

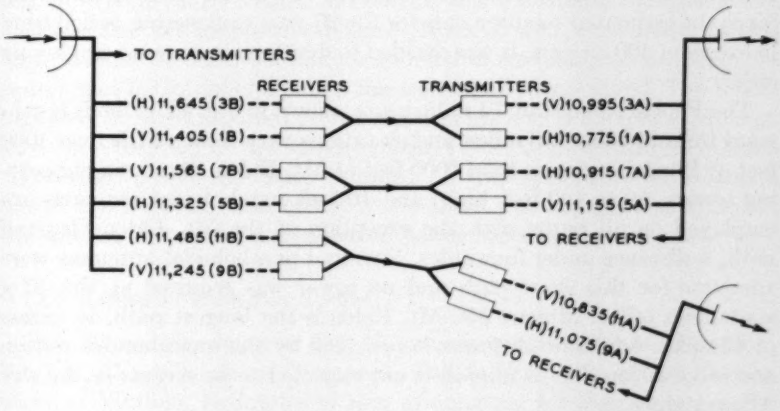


Fig. 43 — Typical frequency assignments at TJ branching point.

There will be instances where it will become desirable to add a branching route to a fully loaded route. To do this, some radio channel frequencies must be reused at the branching point. In these cases, careful consideration must be given to near-end and far-end crosstalk. A requirement of 50 db must be placed on the reduction of any undesired signal arising from these exposures. This precludes the use of periscope antennas. In order to meet this requirement, paraboloidal antennas will have to be used at the branching point on those channels having the same frequency. In addition, frequencies should be selected so that the channels common to both routes have the largest angular separation possible.

7.2 *Typical TJ Installations*

7.2.1 *Phoenix-Flagstaff System*

Prior to the installation of the TJ radio system between Phoenix and Flagstaff, Arizona, the facilities along this route consisted of an open-wire line equipped with various types of carrier. By 1956 these facilities had been developed to nearly their maximum capacity. New circuit requirements for northern Arizona would necessitate either an expansion of the old route or construction of a new route. The age and condition of the open-wire route made it uneconomical to rebuild this facility. Cost studies indicated that construction of a new wire-line facility would cost less initially, but that for a cross section greater than 60 two-way circuits, radio multiplexed with L-carrier would be more economical. Since the estimated requirements for the 20-year engineering period were in excess of 400 circuits, it was decided to develop a new route employing radio.

The Phoenix-Flagstaff TJ radio route, shown in Fig. 44, extends northward from Phoenix 133 miles, and elevations vary from a little over 1000 feet at Phoenix to more than 9000 feet at Mt. Elden. Short self-supporting towers, 60 to 120 feet high, and 10-foot paraboloidal antennas are employed on all paths with the exception of the Mt. Elden-Flagstaff path, a distance under four miles. Five-foot paraboloidal antennas were adequate for this short path and no tower was required at the Flag staff main office. Mingus Mt.-Mt. Elden is the longest path, in excess of 47 miles. Adequate clearance is provided by the mountainous terrain and attenuation due to rainfall is not expected to be serious in the dry Arizona climate.

It is interesting to note that freight-type containers, 28½ feet long by

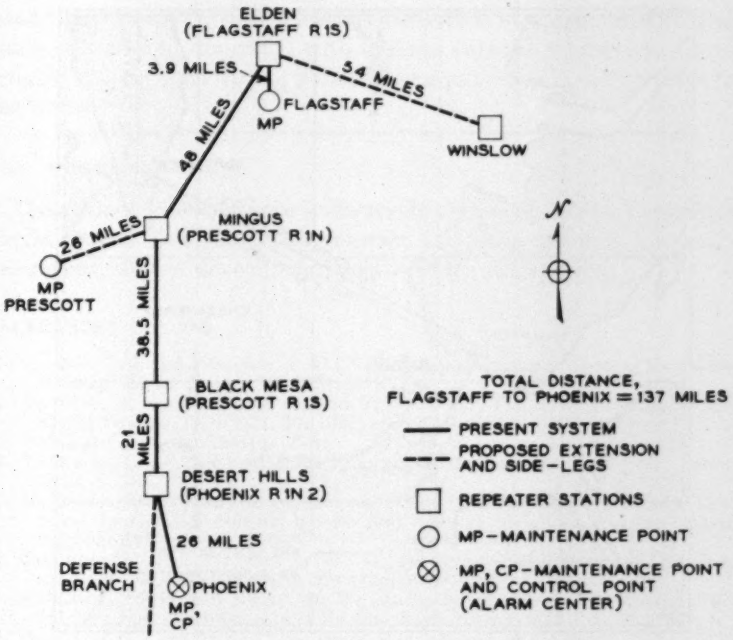


Fig. 44 — Route plan of the Phoenix-Flagstaff TJ system.

8 feet wide, were used for repeater buildings on this route. These buildings were temporarily located in Phoenix, where the radio equipment was installed and a large portion of the equipment tests were made. In this way, most of the time normally consumed in travelling to and from repeater sites for initial equipment line up and tests was saved. The buildings were then transported to the repeater sites as a complete package with the exception of the stand-by power sets. Commercially available "no-break" power sets were installed on this system at all locations except the Phoenix main office, where there was sufficient capacity in the existing L-3 motor alternator to power the additional load.

7.2.2 *Kiptopeke-Hampton System*

This TJ system, shown in Fig. 45, extends northward from Hampton, Virginia, 22.5 miles across Chesapeake Bay to Kiptopeke on the Eastern Shore of Virginia. Estimates of rain attenuation for this section of the country indicate that reliable transmission can be achieved over these

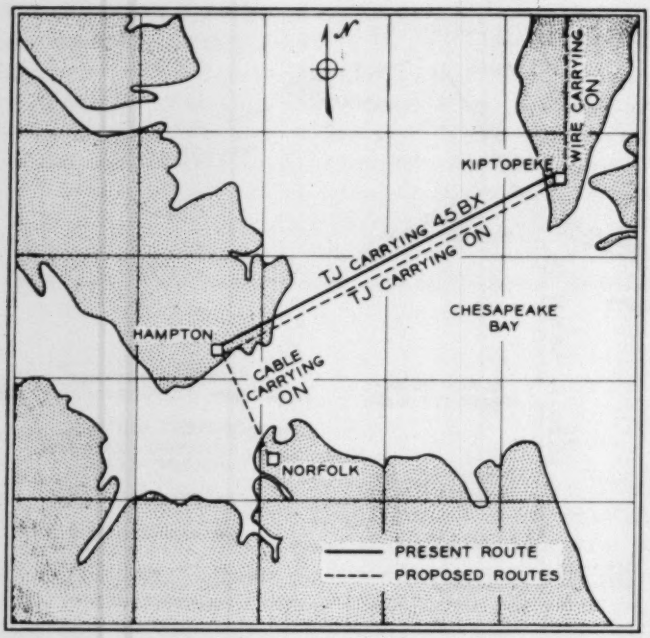


Fig. 45 — Hampton-Kiptopeke TJ system with proposed additions.

distances provided the system has a fading margin of 40 db. However, in engineering a microwave system to operate on a long over-water path, two propagation phenomena besides rain attenuation deserve careful consideration. These are (a) fading due to inverse bending, and (b) reflections from the surface of the water. In the former case, additional tower height is usually provided to assure reliability. In the latter case, some arrangement, such as a high tower at one end and a low tower at the other end of the path, is employed to place the reflection point advantageously so that cancellation will not occur. As can be seen, the remedy for inverse bending is not the remedy for reflective paths.

The Kiptopeke-Hampton TJ system employs periscope antenna systems at both ends of the path. The reflective antenna height at Kiptopeke is 266 feet, and at the Hampton end it is 260 feet. The received signal level is -32 dbm and, with 36 channels of single-sideband suppressed-carrier multiplex, 48 dba at 0 TLP occur at a receiver input of -75 dbm resulting in a fading margin of 43 db. The antenna heights provide grazing clearance when $k = 1/2$. Under normal propagation

conditions, $k = 4/3$, the mid-path clearance is in the order of 15 fresnel zones. Frequency diversity, with 240-mc channel separation, provides reliable transmission during periods when reflection is taking place from the water.

VIII. ACKNOWLEDGMENT

The system described here is the result of the efforts of several departments of Bell Telephone Laboratories, including research, systems development, device development and systems engineering.

REFERENCES

1. Grieser, T. J. and Peterson, A. C., A Broadband Transcontinental Radio Relay System, Elect. Engg., **70**, 1951, p. 810.
2. Roetken, A. A., Smith, K. D. and Friis, R. W., The TD-2 Microwave Radio Relay System, B.S.T.J., **30**, 1951, p. 1041.
3. Bell Lab. Rec., **24**, 1946, p. 175.
4. Tillotson, L. C., A Short-Haul Microwave Transmitter, Bell Lab. Rec., **33**, 1955, p. 131.
5. McDavitt, M. B., 6000-Megacycle-per-Second Radio Relay System for Broadband Long-Haul Service in the Bell System, A.I.E.E., General Meeting, October 1957, paper 57-1044.
6. Cronburg, C. I. L., Jr. and Aruck, M., 96-Channel Multiplex for ON Carrier on Radio, A.I.E.E., Seattle Meeting, June 1959.
7. Hathaway, S. D. and Evans, H. W., Radio Attenuation at 11 kmc and Some Implications Affecting Relay System Engineering, B.S.T.J., **38**, 1959, p. 73.
8. Pound, R. V., Electronic Frequency Stabilization of a Microwave Oscillator, Rev. Sci. Inst., **17**, 1946, p. 490.
9. Pomeroy, A. F., to be published.
10. Lewis, W. D. and Tillotson, L. C., Non-Reflecting Branching Filter for Microwaves, B.S.T.J., **27**, 1948, p. 83.
11. Small, R. H., to be published.
12. Dahlbom, C. A., A Transistorized Signaling System, Bell Lab. Rec., **37**, 1959, p. 254.
13. Jakes, W. C., Jr., A Theoretical Study of an Antenna-Reflector Problem, Proc. I.R.E., **41**, 1953, p. 272.
14. Ohm, E. A., A Broadband Microwave Circulator, I.R.E. Trans., **MTT-4**, 1956, p. 210.

Evaluation of Message Circuit Noise

By A. J. AIKENS and D. A. LEWINSKI

(Manuscript received April 5, 1960)

The reduction and control of message circuit noise requires that its effects be quantitatively characterized. This is achieved by measuring some physical attribute of the electrical noise such that two noises that are judged to be equally interfering are assigned approximately equal numerical magnitudes. To give meaning to such magnitudes, the scale of measurement is related by way of subjective assessment into terms useful to telephone engineering. Pertinent to message circuit noise measurement is the characterization of the relative interfering effect of single-frequency noise components and the way the ear combines these components to indicate the total effect. The correlation between a noise measurement and its associated over-all end-effect may be made by finding the transmission-loss equivalent of the noise being measured or by the direct application of telephone-user opinions expressing degree of transmission satisfaction. In addition to these topics, this paper discusses certain engineering aspects of noise evaluation.

I. INTRODUCTION

From the time of the first commercial telephone system, transmission engineers have been concerned with the problem of message circuit noise evaluation. Prior to the turn of the century, progress was slow and limited to crude estimates because of the lack of means for performing measurements. One's hearing mechanism had to serve as the indicating meter, and the scale of measurement was simply one's expression of attitude concerning the noise.

Initially, the most accurate determination of noise magnitudes was through frequency analysis and computation. A given noise voltage was first analyzed into its spectrum, which was then weighted according to a predetermined curve that characterized the relative interfering effect of frequencies in the telephone channel. The total noise magnitude was then obtained by summing the weighted components on a power basis. A quicker but less accurate way was to make an over-all noise measurement. "Circuit noise" was simply compared by ear to a variable-ampli-

tude tuned buzzer, but due to the lack of objectivity the measured magnitudes often tended to disagree because of differences between operators.

With the advent of electronic measuring devices the picture began to change. Vacuum-tube voltmeters, with square-law detectors and associated circuitry to simulate weighting characteristics, would assign approximately the same magnitudes to noises producing the same order of interference. This was a significant advance, since it enabled the introduction of uniform yardsticks to give value-type interpretations of the severity of interference associated with each measured magnitude. The original yardstick was based on the concept of "noise transmission impairment," and subjective tests were performed to relate any given noise magnitude to an equivalent transmission loss.

Since that time, which dates back to the early 1930's, electronic noise measuring sets, their use, and the interpretation of measured magnitudes have undergone change. In addition to these changes, the operational nature of noise evaluation requires that many of the techniques and ramifications be constantly re-examined. It is not surprising, therefore, that the following questions are often asked:

i. How does *this* noise set measure a given noise?
ii. What are the units? Must a correction factor be added?
iii. Where is the best place to make the measurement?
iv. What type of measurement should be made?
v. What does the measurement mean operationally? What is considered a tolerable magnitude?

While much has been written on the subject, a large part of noise practice exists only in specialized reports on specific problems. To relate this knowledge and experience is, of course, beyond the scope of this article. Our purpose, rather, is to present a tutorial treatment of the most important historical, theoretical and practical aspects of message circuit noise evaluation.

We begin by discussing the fundamental problem: the measurement of circuit noise such that noises having equal "interfering effect" are assigned equal magnitudes. Pertinent are the concept of frequency weighting (i.e., the characterization of noise effects versus frequency); the manner in which weighted components combine; and the units of noise measurement. Background is provided on the choices over the years of reference for noise measurement and the effects of these references on numerical noise magnitude.

Next, the discussion follows the historical development of noise measuring devices, starting with a review of the features of the earliest device and ending with the 3A Noise Measuring Set.*

Highlights of the quantitative aspects of noise evaluation are then reviewed. Emphasis is placed on noise originating outside the telephone system, the application of the "telephone influence factor" to predict power system noise influence, and the use of noise measuring sets to estimate longitudinal and shunt unbalances in open-wire lines and cable pairs, assuming that external induction exists.

Finally, we consider the assessment of noise magnitude from the subscriber's standpoint. The discussion is centered on the use and relative merit of noise transmission impairments and "grade of service."

II. BASIC CONSIDERATIONS IN NOISE MEASUREMENT

2.1 *General*

An electrical disturbance in a telephone circuit — that is, any signal that does not convey intelligence — appears acoustically as *noise* at the output of the receiver of the telephone instrument. As noise capable of detection by the ear, it may be barely audible or it may be obtrusive enough to impair the reception of a telephone message. Noise is "high" when it is bothersome; it is "low" when its effects are insignificant. Thus, the evaluation of message circuit noise must concern itself with a quantification of noise based on the subjective reactions experienced by telephone users.

To achieve this objective, it has been found appropriate to perform these two operations:

(a) measure some physical attribute of the electrical noise such that two noises that are judged to be equally interfering are assigned approximately the same numerical magnitude; and

(b) subjectively assess the severity of the interference associated with each magnitude on the resulting physical scale, and relate the assessment into acceptability criteria such that the scale of measurement becomes useful to telephone engineering.

These two steps broadly define the underlying principles of message circuit noise evaluation: the first prescribes the way noise is to be meas-

* The 3A set is covered in detail in a companion paper.[1]

ured; the second provides the information necessary to the interpretation of the measurement.

In general, however, conversation over a telephone may be rendered more difficult not only by the presence of circuit noise, but also by the presence of room noise in the location of the telephone, the reaction of a particular subscriber, the way he uses the telephone and the volume level of the telephone conversation itself. Although it is impractical to incorporate accurately all of these parameters, the principle of measurement and the method for the subjective assessment of magnitudes allow, to a good approximation, for these additional effects.

From the measurement standpoint, it is sufficient to assign equal noise magnitudes on the basis of *moderately bothersome* noises that are judged to be equally interfering in the presence of *average* levels of received speech volume and room noise. Two characterizations are necessary: a "weighting function," which characterizes the relative interfering effect of single-frequency noise voltages, and a law of summation expressing the manner in which the weighted frequencies combine to produce the total effect. The weighting function is derived by evaluating and combining the relative annoyance of moderately loud single frequencies in the absence of speech and the relative impairment of speech by these frequencies, when the speech is at the average received volume level. The desired noise magnitude is obtained by weighting a given noise spectrum in accord with such a weighting characteristic, and then integrating the weighted spectrum by the associated law of summation.

Step (b) enables one to account for the effect of all other conditions of electrical circuit noise, received speech volume and room noise. This requires a multidimensional relationship expressing the degree of satisfaction, or degradation for the range of noise magnitudes (which result from the manner of assigning values as described above) in combination with any given levels of received volume and room noise.

2.2 *Frequency Weighting Characteristics*

Speech is impaired by the coincident presence of one or more foreign sounds. The mechanism is such that while one is listening to one sound the ability to listen to another is reduced — the degree of reduction being dependent on the relative amplitude of the two sounds. Quantitatively, sounds of frequencies below 900 cycles have a tendency to impair their adjacent higher frequencies rather than their adjacent lower frequencies. Above 900 cycles this effect is not as pronounced. However,

maximum impairment always occurs near the frequency of the impairing sound.

Annoyance, on the other hand, refers to the subjective discomfort caused by noise during the absence of speech. It may be considered from the standpoint of loudness. As the volume level of a sound is increased, the louder and thus more annoying it becomes, although the effect is not directly proportional.[2] At low volumes, a high-frequency sound tends to be more annoying than a low-frequency sound, whereas at high volumes all frequencies tend to become equally annoying.

Theoretically, therefore, a particular weighting will characterize the effects of noise versus frequency only at specific volume levels. From a practical standpoint this is not a serious drawback. Consider, for example, the Fletcher-Munson loudness contours[3] shown in Fig. 1. Notice that, if the various noise components per cycle are confined to an approximate 25-db range, then the relative loudness does not change appreciably. Likewise, the change in the impairment mechanism of single frequencies is not appreciable over this range. Because of rigid noise

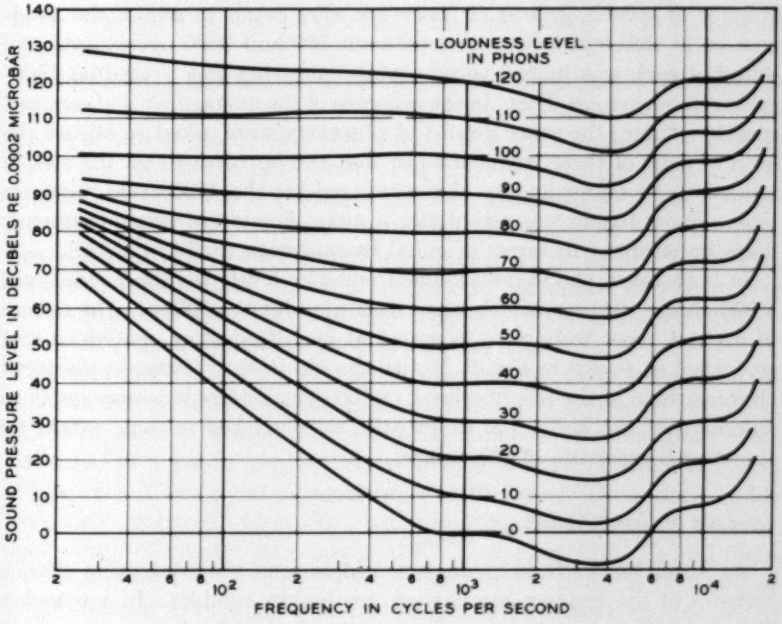


Fig. 1 — Free-field Fletcher-Munson equal-loudness contours for pure tones.

control throughout the Bell System plant, the 25-db range is usually not exceeded. Hence, it is the practice to use only one weighting characteristic for all applications, where such a characteristic is determined relative to some moderately bothersome level encountered in practice.

Throughout the years, weighting characteristics have changed with the gradual improvement of the telephone set. The first standard weighting characteristic used in an electronic measuring set, called the 144-line weighting, related to the deskstand telephone set prevalent during the 20's and early 30's. The second, called the F1A-line weighting, related to the 302 set and the third, called the C-message weighting, relates to the present 500 set. Although each of these was determined by a different subjective evaluation technique, they all incorporate the impairment and annoyance effects of noise for average conditions of message circuit telephony.

Since the determinations of the 144 and F1A weightings are well documented,[4,5] it is sufficient to describe briefly the technique for the derivation of the present C-message weighting. Two tests were conducted, one in the absence and one in the presence of speech. In the absence of speech, groups of observers were asked to adjust the loudness of 14 different frequencies between 180 and 3500 cycles until the sound of each was judged to be equally annoying as a -59-dbm 1000-cycle reference tone. Then, in the presence of the average level of received speech volume, the same groups of observers were asked to adjust the level of each of these frequencies so that the impairment on the speech tended to be equivalent to that produced by the 1000-cycle reference tone. The outcome was essentially a quantification of single-frequency noise impairment in terms of equal transmission quality.

In both cases, the various single-frequency levels were measured into a 900-ohm resistance, which was substituted for the 500 set. The results of each of these tests were averaged at each frequency, combined and smoothed as shown in Fig. 2. Notice the dip in the relative impairment characteristic in the neighborhood of 500 cycles. Here it is reasonable to assume that the individual frequencies were masked to some extent by the speech spectrum of the talker.

2.3 *Law of Combination*

As is well known from loudness considerations, the integrating characteristics of the hearing mechanism are highly complex. In particular, the mode of integration is amplitude-dependent and influenced by masking. It has been found, however, that weighted noise components are added approximately as power, provided the noise is confined to rela-

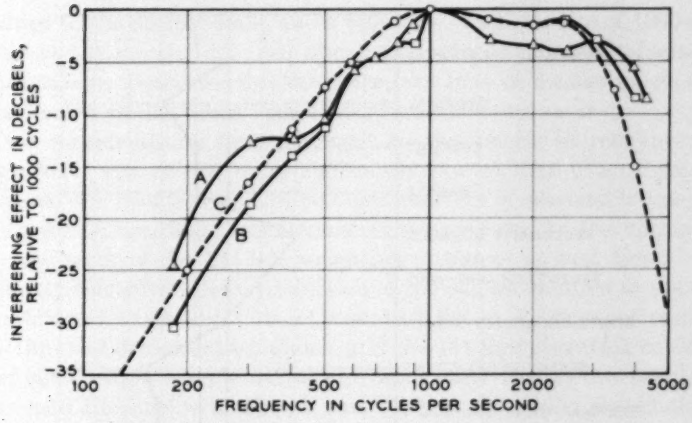


Fig. 2 — Interfering (subjective) effect in decibels of various single frequencies relative to the frequency (approximately 1000 cycles) of maximum interfering effect: A, interference due to impairment; B, interference due to annoyance; C, adopted combined effect.

tively low levels. Since this is usually the case, power addition is most appropriate.

While power addition is consistent with Fletcher's law of loudness addition for low-level weighted signals, early articulation studies of the effects of two or more noises also showed that the combined noise effect adds on a root-sum-square basis.[6] The latter was further borne out by subjective correlation of a number of noise measurements made with the F1A-line weighting characteristic using square-law detection.[5] Thirteen different types of noise that were judged to be equally interfering measured substantially equal in magnitude, as shown in Fig. 3. For comparison, the particular F1A-line weighted power measurements are indicated in decibels relative to open-wire line noise.

Further justification for the use of power addition was recently established in tests to determine the applicability of the C-message weighting. Different pairs of moderately interfering bands of thermal noise were presented to a group of observers who were asked to change the level of one until it was equal in interference to the other. These same bands were also compared objectively, by computing the differences between their weighted spectra on a power ($10^{db/10}$), voltage ($10^{db/20}$) and high-level loudness ($10^{db/40}$) basis.* The average subjective decibel differences in interference were found to agree remarkably well with the computed

* These three modes of summation are good approximations to Fletcher's law of loudness addition for low-, medium- and high-level weighted signals heard over the telephone.

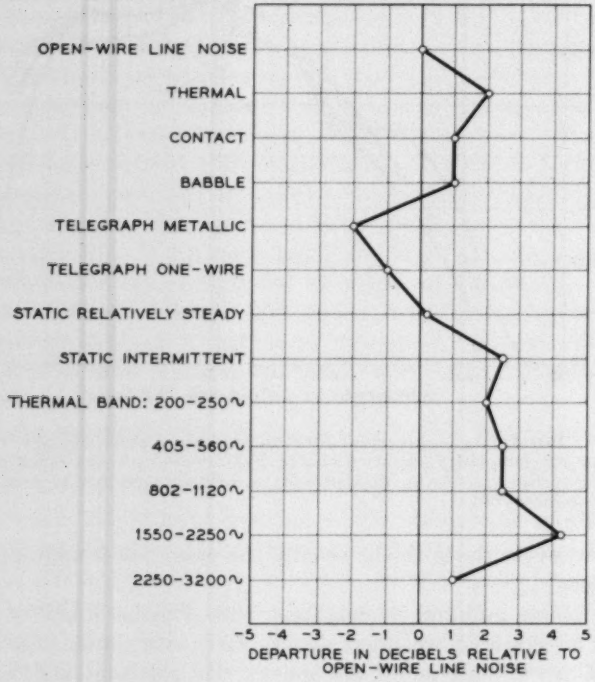


Fig. 3 — Difference between F1A-line weighted noise measurements for noises judged to be equally interfering.

differences based on power addition, whereas correlation for voltage and high-level loudness addition was poor.

2.4 *Units of Noise Measurement and Noise Magnitudes*

While a weighting characteristic and its companion law for the summation of weighted components essentially prescribe the way message circuit noise is to be measured, one must also be provided with a noise reference and a scale of measurement. At the time of the ear-balance (i.e., acoustic comparison) method of noise measurement, the unit of measurement was simply the "noise unit." It was defined to be equal to 10^{-6} of the current output of a 240-cycle buzzer, which served as the comparison standard. The value of current above this amount was numerically equal to the number of noise units.

With the advent of electronic noise measuring sets, the scale was

changed to the decibel scale, and a reference of 10^{-12} watts of 1000-cycle power at the input (i.e., −90 dbm) was proposed for the reference of measurement. Based on this reference, the unit of measurement came to be known as the "dbrn" (decibels above reference noise).

Two considerations were pertinent to this choice of reference: the power level was chosen to be sufficiently low so that when noise was measured the results would be positive numbers of reasonable size; and the choice of the 1000-cycle frequency corresponded to the normalized zero loss point of the 144-line weighting prevalent at that time.

Shortly thereafter, the introduction of the 302 set resulted in the need for a different weighting. This set was much flatter in frequency response than the older deskstand telephone, and the 144-line weighting could not be applied. After completion of the work that led to the previously mentioned F1A-line weighting, it was felt that something should also be done about the reference of measurement. The newer line weighting admitted a substantially wider band of frequencies and for the same reference the two weightings would measure a given noise numerically different. Tests showed that with the newer weighting the result was about 5 db higher for several types of noise, including 3 kc thermal noise.

To bridge this gap, a new reference was proposed in conjunction with the newer weighting characteristic. This reference (i.e., zero on the meter) was a 5-db upward adjustment, from 10^{-12} to 3.16×10^{-12} watts of 1000-cycle power, or from −90 dbm to the 5-db-less-sensitive −85 dbm. Hence, for a uniform 3-kc band of thermal noise one would now obtain the same numerical magnitude of measurement with either weighting. On the other hand, while the old reference required 10^{-12} watts of 1000-cycle power to give a zero indication on the meter, 3.16×10^{-12} watts, or 5 db more of 1000-cycle power, would be needed to give a zero indication with the new reference. To distinguish between the two references, the newer one was designated "dbrn adjusted" or "dba."

At the present time, the picture has further changed. The introduction of the 3A Noise Measuring Set with its new C-message weighting posed a similar problem regarding its reference for measurement. This weighting, as will be seen, has a wider equivalent spectrum than the F1A-line weighting. To follow the approach described above would require a still higher equivalent reference in order to obtain the same numerical measurement for 3 kc thermal noise. Although this step was considered, it was deemed unwise for with a higher (less sensitive) equivalent reference, many of today's low noise levels would be assigned "negative" magnitudes.

To circumvent this possibility, it was decided to return to the basic 10^{-12} watt reference, that is, to -90 dbm. This reference, although numerically equal to the original dbrn at 1000 cycles, is not its equivalent when viewed in terms of weighted noise measurements. While this latest reference is called dbrn, in actual fact it is "dbrn C-message" to distinguish it conceptually from "dbrn 144-line." For instance, as will be shown later, 3 kc band-limited random noise will read 6.5 db higher with C-message weighting than with 144-line weighting. In the case of an F1A-line weighted noise measurement, both the weighting characteristic and the different reference (dba versus dbrn) will affect the results.

The relationships can best be illustrated with some examples. The three different weighting characteristics are shown in Fig. 4. Since these curves reflect relative subjective noise effect (or relative interfering effect), the maximum of each curve is normalized at 0 db. Note that the maximum of each curve occurs at approximately 1000 cycles.

Consider now the measurements that would be obtained for two extreme cases of noise: (a) a 1000-cycle noise voltage and (b) a random noise voltage. The former is relatively easy. We need only compare the given signal to the reference of measurement, since a 1000-cycle signal suffers no weighting. For illustration, let us assume a 1000-cycle voltage such that the noise power is 10^{-3} watts at the input of a noise set, capable of each of the three measurements. We would obtain the following

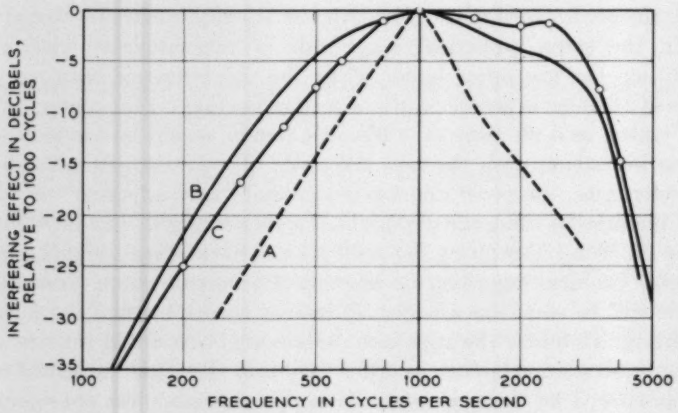


Fig. 4 — Response vs. frequency of: A, 144-line weighting; B, F1A-line weighting; C, C-message weighting.

noise levels:

$$N_{\text{dbrn(144)}} = 10 \log_{10}\left(\frac{10^{-3}}{10^{-12}}\right) = 90 \text{ dbrn},$$

$$N_{\text{dba}} = 10 \log_{10}\left(\frac{10^{-3}}{3.16 \times 10^{-12}}\right) = 85 \text{ dba},$$

$$N_{\text{dbrn (C-message)}} = 10 \log_{10}\left(\frac{10^{-3}}{10^{-12}}\right) = 90 \text{ dbrn}.$$

Now consider the application of the three weighting characteristics to a random noise voltage whose power spectrum $p(f)$ is rectangular and confined to a 3-kc band. To determine the three noise measurements, it is sufficient to compare the total unweighted power P to each of the three weighted powers $P_i'(i = 1, 2, 3)$ and express these differences on a decibel basis.

For the assumed power spectrum, $p(f)/3000$ is constant per cycle, say p, at the input to our hypothetical noise set. Thus the weighted powers P_i' are given by

$$P_i' = p \int_0^{3\text{kc}} w_i(f)\, df,$$

where $w(f)$ is a particular weighting characteristic, expressed as $10^{\text{db}(f)/10}$ to comply with the power addition requirement for weighted noise components. In this notation, the values db(f) are simply the relative subjective effects provided by each of the three weighting curves on Fig. 2. Thus the reduction in input power suffered by weighting is given by

$$10 \log_{10}\left(\frac{P}{P_i'}\right) = 10 \log_{10}\left[\frac{3000}{\int_0^{3\text{kc}} w_i(f)\, df}\right].$$

The problem of solving the integrals $\int_0^{3\text{kc}} w_i(f)\, df$ is readily facilitated by numerical approximation. Fig. 5 shows the three weighting curves transformed from relative decibels to numerics, $w(f) = 10^{\text{db}(f)/10}$.

Since $w_i(f)\,|_{f=1000} = 1$, the area under the line $w(f) = 1$ from 0 to 3 kc must equal 3000 square units. Hence, if we let this area be A and the areas under $w_i(f) = A_i$, then the reduction in power due to weighting is:

$$10 \log_{10}\left(\frac{A}{A_i}\right).$$

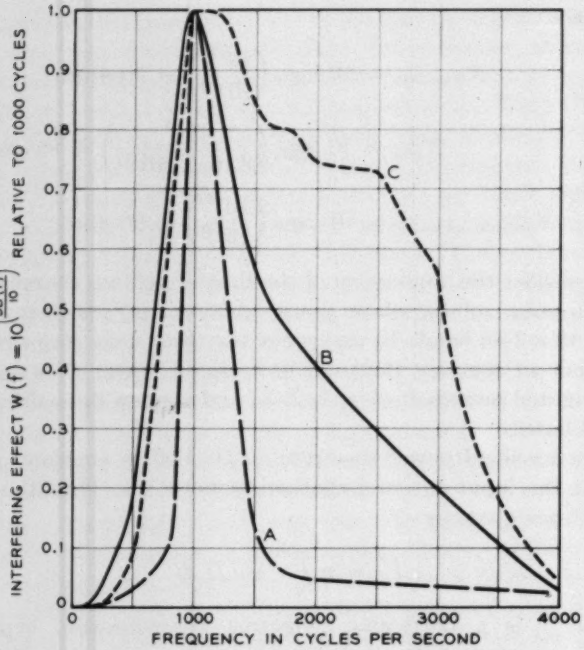


Fig. 5 — Result of transforming the three weighting characteristics shown in Fig. 4 from relative db to $10^{db/10}$: A, 144-line weighting; B, F1A-line weighting; C, C-message weighting.

It is found approximately that

$$A_1 = A_{144} \doteq \frac{A}{6.2},$$

$$A_2 = A_{\text{F1A}} \doteq \frac{A}{2}.$$

$$A_3 = A_{\text{C-message}} \doteq \frac{A}{1.4}.$$

Hence the effect of the three weightings is to reduce a flat 3-kc spectrum by approximately 8, 3 and 1.5 db respectively. For example, given a flat 3 kc-spectrum that measures 10^{-3} watts at the input to our hypothetical noise set, the three respective noise levels are found to be:

$$N_{\text{dbrn(144)}} \doteq 10 \log_{10}\left(\frac{10^{-3}}{10^{-12}}\right) - 8 = 82 \text{ dbrn},$$

$$N_{\text{dba(F1A)}} \doteq 10 \log_{10}\left(\frac{10^{-3}}{3.16 \times 10^{-12}}\right) - 3 = 82 \text{ dba},$$

$$N_{\text{dbrn(C-message)}} \doteq 10 \log_{10}\left(\frac{10^{-3}}{10^{-12}}\right) - 1.5 = 88.5 \text{ dbrn}.$$

Evidently for other complex noise voltages these relationships are not valid. In general, given any noise $g(t)$, whose long-time average power spectrum is $p(f)$ (where $f_1 \leqq f \leqq f_2$), then the noise level of $g(t)$ is:

$$N_{\text{db(ref)}} = 10 \log_{10}\left[\frac{\int_{f_1}^{f_2} p(f)\ 10^{\text{db}(f)/10}\, df}{\text{reference power}}\right].$$

III. NOISE MEASURING DEVICES

3.1 *Early Devices*

It has been shown that a noise weighting characteristic, a companion law for the summation of weighted components, a reference level and a scale of measurement are all necessary to effect a meaningful measure of circuit noise. It will now be discussed how these basic requirements are, and have been, simulated in instruments to obtain the measurements directly.

The first apparatus for the measurement of circuit noise was introduced shortly after 1900. It consisted of a sort of acoustic comparator wherein the weighting and summation was achieved directly by ear. One first listened to the acoustical output of a telephone receiver connected to a "noisy" telephone line and then, by means of a switch, listened in the same receiver to a small inductor alternator generating a 240-cycle voltage. The intensity of the sound produced by the output of the altenator was adjusted by means of a potentiometer until the observer felt that the "interfering effect" of the alternator tone was the same as the "interfering effect" of the noise. The result of the measurement was expressed in microamperes in the telephone receiver. It may be observed that, in using this device, each "noise measurer" became an individual interpreter of the subjective-objective relation of noise.

Subsequently, the alternator was replaced by a battery-operated buzzer, which supplied a constant volume source of tone called a "noise

standard." An additional improvement was made by providing the potentiometer with a numerical scale, graduated in "noise units."[7] This device, called the 1A Noise Measuring Set (see Fig. 6) did not, however, attain any greater objectivity than its predecessor. It remained on the scene until 1929, when it was superseded by the 1A Noise Amplifier (a device intended for noise measurement on toll circuits). The latter was the first to incorporate frequency weighting, and it employed the 6A Transmission Measuring Set as an indicating device. The elimination of the "noise measurer" as part of the measuring system was of considerable importance in that it eliminated most of the difference between operators.

3.2 *Electronic Noise Measuring Sets*

In accord with the principle of message circuit noise measurement, an ideal noise measuring device must simulate a procedure that will assure that two noises judged to be equally interfering are assigned the same numerical magnitude.

Thus, ideally, for any given noise voltage $g(t)$ the device must first separate $g(t)$ into its frequency components. This might be achieved by a series of bandpass filters, each with passband $\Delta\omega$. The device must then sense the variations in amplitude at the output of each filter to gauge the relative interfering effect of all components present. Then, by means of variable attenuation, each component must be weighted by an amount equal to its relative effect as determined by the previous operation. Next, the device must transform and sum the weighted amplitude components by a rule to indicate the total interfering effect of the original noise voltage $g(t)$. Finally, the outcome should be time-averaged over a period comparable to the hearing response and the resulting magnitude suitably indicated on a numerical scale as a function of time.

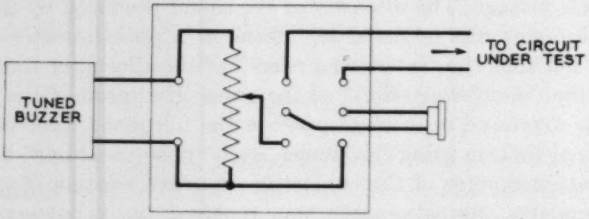


Fig. 6 — Schematic of 1A Noise Measuring Set. Potentiometer was adjusted until buzzer tone appeared to be equally interfering as noise on circuit under test.

From a practical standpoint, this simulation can be simplified as discussed in Section 2.1. The weighting process is simply achieved by a suitable network, the response of which approximates the interfering effect for the nominal range of noise levels encountered. For the transformation and summing process it is sufficient to use square-law rectification, which is appropriate for the noise levels of interest. Lastly, the rectified output, which is then proportional to the rms value of the weighted noise voltage, may be averaged over a specified period of hearing response and indicated on a dc meter that is calibrated in terms of input to the set in decibels from the reference of measurement.

Thus, operationally, all electronic noise measuring sets ought to behave in a like manner. The essential differences will be only in the type of weighting and the particular reference of measurement that appears to be appropriate, as well as in specific features that are dictated by the telephone system environment and special noise measuring needs. The latter refer mainly to the input arrangements and specific weightings, such as might be required in the measurement of program circuit noise or noise across the receiver.

One of the input arrangements of particular importance is the noise-to-ground input. Its inception dates back to the first major activity in noise evaluation: the inductive coordination of power and telephone systems, begun in 1912.[8] The investigators found that balance between the two sides of the telephone circuit at noise frequencies was an all-important factor. The power circuit induces comparatively high voltages to ground on each wire of the telephone pair. Unless the pair is well balanced, either longitudinally or to ground, these induced voltages will cause metallic currents in the transmission circuit. Even slight unbalances will result in such currents, which appear as metallic noise voltages across the terminals of the circuit. While it was realized that excessive metallic noise voltages of this type could be minimized by good maintenance and design, the means of guiding such mitigation required a noise set that would measure both "noise-metallic" and "noise-to-ground" accurately.

Although the ear-balance noise measuring sets were used to measure noise-to-ground (as well as noise-metallic), it was not until the advent of the electronic noise measuring set that these measurements attained the necessary sophistication and objectivity.

3.3 *The 2A — Forerunner of the 2B and 3A*

The 2A Noise Measuring Set,[9] introduced in 1937, was the first device

that adequately realized the primary objective of message circuit noise measurement. It was portable, and went a long way to imitate the important qualities of the hearing mechanism discussed in Section II. A functional diagram of the set is shown schematically in Fig. 7.

This set consisted of various input circuits, a number of filters to simulate weighting characteristics (the basic one for noise measurements being the 144 weighting), an attenuator, a three-stage vacuum tube amplifier, a quasi-rms (copper oxide) detector for power summation of weighted components and a decibel meter with a 200-millisecond integration time* to indicate the noise level in dbrn. Additional features included a self-contained battery power supply, and means for internal calibration. While these components were designated primarily from the standpoint of noise measurement, the set had other capabilities. For example, means were also made available for volume and sound level measurements — two features that were later standardized in other primary measuring devices.

The input impedances were chosen to be compatible with the telephone system environment. A 600-ohm line input was provided for terminating noise metallic measurements on toll circuits. This input was designed to work with either the 144-line weighting or a flatter weighting suitable for the measurement of noise on 8-kc program circuits. In addition, a 2000-ohm bridging impedance was supplied to measure noise across the receiver (the latter being low impedance). The weighting for receiver noise measurements was a modification of the 144-line weighting, wherein the line-to-receiver transfer characteristic of the deskstand telephone was taken into account.

In addition to these three basic input arrangements, a 6000-ohm bridging input was provided to enable the measurement of noise-metallic on working telephone circuits. The input circuit for noise-to-ground consisted of 100,000 ohms in series with the line input. This arrangement presented a high input impedance to ground and reduced

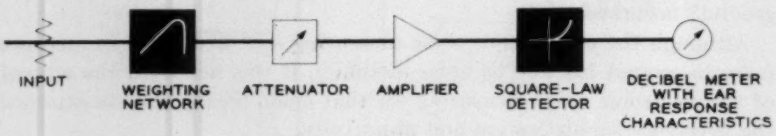


Fig. 7 — Schematic of 2A Noise Measuring Set.

* This is approximately the time required by the ear to appreciate the full loudness of a sound.

the sensitivity to make the indicated noise-to-ground magnitude comparable to the noise-metallic magnitude.

To make any of these measurements one simply connected the circuit under test to the proper input and adjusted the attenuator until the meter pointer gave a scale indication. The measurement was the sum of the attenuator setting and the meter indication. The former had a 60-db range, whereas the meter scale had an 18-db range. Reference noise in all cases was the meter reading (i.e., "zero") obtained with 10^{-12} watts of 1000-cycle power dissipated at the point of measurement. The actual minimum measurable noise was in the order of 10 dbrn.

The "sound input," in conjunction with a suitable microphone and matching transformer, permitted the set to be used for sound level measurements. While the minimum measurable level depended on the microphone and the transformer, the use of the standard 630A condenser microphone with the 111A repeat coil enabled the measurement of sound level as low as 55 db above reference. The weighting used in conjunction with such measurements corresponded closely to the "A" weighting currently found in sound level meters. The reference for sound measurements was chosen equivalent to 10^{-16} watts per square centimeter at 1000 cycles.

A major drawback of this set was that, for one-half of its 12 possible measurements, it was necessary to add various different correction factors to obtain the correct numerical magnitudes.

3.4 *The 2B and 3A Noise Measuring Sets*

In essence, the 2B Noise Measuring Set was a modified 2A. Introduced in 1941, it incorporated the F1A-line and the so-called HA1-receiver weighting for noise measurements in dba, two extra sound weightings, and improved means of internal calibration. The new noise weightings were needed because of the advent of the 302 telephone set, which had a different response characteristic than the earlier deskstand telephone set for which the 2A set with its basic 144-line weighting had been designed.

For simplicity, the two new weightings were obtained by changing the responses of the 144-line and receiver networks. The result was that the modified networks had an inherent loss 12 db greater at 1000 cycles than had had the original 144 networks. Since the reference level for F1A-HA1 weighted measurements was −85 dbm, compared to the reference level of −90 dbm for the two 144 weightings, the net difference in meter reading was 7 db. Unfortunately, this necessitated a

variety of additional correction factors for all F1A-HA1 noise measurements — the basic one being 7 db for the F1A-HA1 line, receiver and noise-to-ground measurements.

In the new 3A this situation no longer exists. This set is a scaled-down version of the 2B, eliminating the volume, sound level, and receiver noise measurement features. It is direct-reading, making the need for corrections unnecessary. In addition, the set is smaller and lighter due to the use of miniature components and transistor circuitry. It is more sensitive than either the 2B or 2A, and has an improved quasi-rms detector enabling bridging, terminating and noise-to-ground measurements to zero dbrn with C-message weighting.

IV. QUANTITATIVE ASPECTS OF NOISE MAGNITUDE

4.1 *General*

In addition to the quantification of noise effects, message circuit noise evaluation also deals with noise studies and the control of noise to reduce its magnitudes to levels that do not unduly interfere with telephone service. The problem of control does, of course, hinge on a set of standards—the setting of which is guided by a correlation between subjective effect and measured magnitude. Once the standards are set, however, the quantitative interpretation necessary to noise control becomes a separate field of study. While it is not the purpose of this paper to include the setting of noise standards nor to discuss the engineering steps of controlling noise, it suffices to discuss some of the uses of noise sets that bear directly on noise studies.

4.2 *General Use of Message Circuit Noise Measuring Sets*

Whenever noise is measured on a telephone circuit, one should know the average speech volume level (vu) at the point of measurement. Lacking a specific measurement of vu, one should know the transmission level at the point of measurement from which speech volume may be inferred. Generally, in the case of long distance circuits, one refers the noise measurement to a 0-db transmission level (TL) point.

Since the noise measuring set is an audio-frequency device, noise may be measured at any accessible point in a voice frequency circuit. The circuit may be taken out of service and the measurement made with the line input of the noise set as a termination, or the measurement may be

made on a bridging basis, where regular telephone equipment terminates the line that is idle at the time of measurement.

Noise measurements on carrier or radio channels must, of course, be made at the telephone circuit terminals where these channels operate at voice frequencies.

Last but not least, it is important to monitor each noise measurement. This is necessary to insure that noise is actually being measured and also to observe the character of the noise.

4.3 *Use of Noise Measuring Set To Estimate Telephone Influence Factor*

From the point of view of exchange plant telephony, an important external source of noise is that originating in power systems. This noise is important because, in general, it cannot be reduced to tolerable levels solely by transmission design (unless it is possible to separate power and telephone lines adequately). Because of this situation, the reduction and control of power line noise is and has been accomplished by joint engineering coordination with the power companies.

The most expedient approach is to attempt a solution prior to the actual construction of either a power or a communication facility. This requires an estimation of the amount of noise to be expected from the specified condition of the exposure prior to construction. A method that has been of great help is based on the concept of "telephone influence factor" (TIF).[10,11] This refers to either a voltage or current wave in an electrical supply system, and it is defined as the ratio of the square root of the sum of the squares of "weighted" rms values of all sine wave components to the rms value of the entire wave.

Thus

$$\text{TIF} = \frac{\sqrt{\sum_{i=1}^{n} (A_i W_i)^2}}{A},$$

where A_i is the rms value of the (voltage or current) sine wave component at frequency f_i, W_i is the TIF weighting factor at f_i and A is the rms value of the entire wave.

The TIF weighting function $W(f)$ accounts for the relative interfering effect of single-frequency tones $w(f)$ in the message circuit (e.g., the 144-line, F1A-line or C-message weighting) and the variation with frequency of the inductive coupling between power and communication circuits. The new curve based on the C-message weighting is shown in

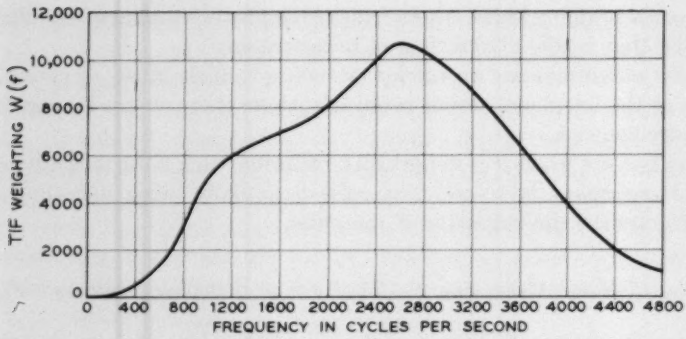


Fig. 8 — Proposed new TIF weighting curve, $W(f) = Kw(f)M(f)$, where K is such that $W(1000 \text{ cycles}) = 5000$; $w(f) = 10^{db(f)/10}$ is the C-message weighting and $M(f) = f/1000$ to approximate the coupling distortion between power and telephone circuits.

Fig. 8. Note that the weighting is given in numerics, with W (1000 cycles) chosen to be 5000.

The measurement of either voltage or current TIF is facilitated by using a noise measuring set (to give the subjectively weighted root mean square magnitudes) in conjunction with a suitable coupler (to provide for the effect of coupling versus frequency between power and communication facilities).[12] A schematic set-up for the 3A Noise Measuring Set is shown in Fig. 9.

Using the voltage coupler, the measurement will be proportional to the $Kv \cdot T$ product (the numerator of the voltage TIF expression):

$$\sqrt{\sum_{i=1}^{n} (V_i W_i)^2}$$

in dbrn, where V_i is the rms value of each sine wave component and W_i the corresponding TIF weighting factor (the latter resulting from the characteristics of the coupler and the C-message weighting). Simi-

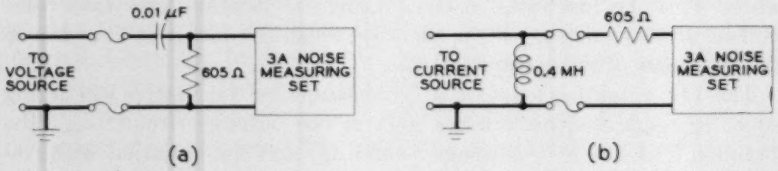


Fig. 9 — Set-up for (a) voltage and (b) current TIF measurements using new 3A Noise Measurement Set.

larly, using the current coupler, the measurement in dbrn will be proportional to the $I \cdot T$ product:

$$\sqrt{\sum_{i=1}^{n} (I_i W_i)^2}.$$

The relation between these measurements and the numerical $Kv \cdot T$ and $I \cdot T$ products is given by

$$Kv \cdot T = 10^{[N_v(\text{dbrn})-43.8]/20},$$
$$I \cdot T = 10^{[N_I(\text{dbrn})-20.2]/20},$$

where N_I and N_v are the corresponding 3A set measurements. Division of these numerical factors by the total rms value of voltage (in kilovolts) or current (in amperes) gives the over-all voltage or current TIF.

In early applications, these factors were used to estimate the noise induced in a given exposure. Today, the greater interest lies in assuring that ac and dc machinery (as manufactured) meet a certain conservative value of TIF.

4.4 *Other Special Uses of Noise Measuring Sets*

As mentioned in Section 3.2, there are two types of unbalances in wire systems that may permit induced "voltage-to-ground" to cause excessively large metallic noise voltages in the transmission circuit. These unbalances, called series and shunt unbalance are shown in their most elementary forms in Figs. 10(a) and 10(b).

In the case of Fig. 10(a), it is assumed that the presence of Z_{sh} does

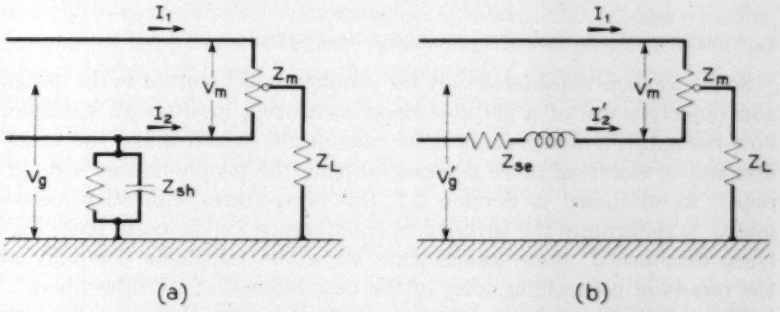


Fig. 10 — Equivalent circuit representations for (a) shunt and (b) series unbalance. For shunt unbalance, 20 $\log_{10}(V_m/V_g)$ = 20 $\log_{10}(Z_m/4Z_{sh})$ = balance ratio in decibels; for series unbalance, 20 $\log_{10}(V'_m/V'_g)$ = 20 $\log_{10}(Z_m/4Z_L)$ = balance ratio in decibels.

not affect the magnitude of V_g and, in the case of Fig. 10(b), it is assumed that the external source of induction is to the left of Z_{ae} and that the Z_{ae} is small compared to Z_L. Both unbalances are assumed to be single entities, and each is therefore an outstanding source of unbalance. Effects of phase shift and attenuation are not taken into account in this example.

A relative measure of these unbalances may be obtained by making noise-to-ground (N_g) and noise-metallic (N_m) noise measurements. For the 3A Noise Set, the standard N_g input arrangement is such that a N_g measurement is 40 db less sensitive than a N_m measurement.

Thus, if one measures both N_g and N_m, the balance ratio is given by

$$N_m - (N_g + 40) = \text{balance in db.}^*$$

In the voice-frequency range, excellent balance performance is represented by ratios < -60 db; good by these between -60 and -50 db, fair by those from -50 to -40 db, and poor by ratios > -40 db.

Another quantity of interest is the severity of the induction itself. This is found by simply measuring N_g. For the new 3A Noise Measuring Set, values of $N_g > 40$ dbrn represent high noise influence, those from 20 to 40 dbrn represent medium noise influence, and those less than 20 dbrn represent low noise influence.

In addition to the above, the N_g input is also useful for measuring low-frequency voltages. These include both 25- and 60-cycle voltages, which are usually not audible. In estimating these components, it is necessary to measure first the weighted and then the unweighted (flat) noise with the appropriate networks. The difference between these measurements will give some indication of the extent of low-frequency induction.

V. SUBJECTIVE ASSESSMENT OF NOISE MAGNITUDE

So far we have considered only the ramifications involved in the design and objective use of a suitable noise measuring instrument. Consider now the subjective aspects of noise magnitude, which make the measurement of electrical noise meaningful from the telephone user's standpoint. As discussed in Section 2.1, this necessitates subjective assessments to determine the severity of interference versus noise level N_m. Once established, these assessments serve the telephone company as the means of controlling noise in the best interest of its subscribers.

The results of two types of assessments are currently in use. One set of results *measures* noise magnitude in terms of "equated transmission

* In the absence of an induced voltage V_g, this method will not yield a meaningful result. However, in such cases the existence of unbalances will be no problem.

loss"; the other provides expressions of attitude to noise as a function of magnitude in the presence of different amounts of received speech volume. Since these results are completely different in concept, they will be treated separately.

The first dates back to an early philosophy, which postulated that all factors that degrade a telephone conversation and make it poorer than face-to-face conversation, e.g., noise, distortion, sidetone and volume loss, should be expressed in similar units. As such, consideration could be given to the summation of the effects of these factors to produce a figure of merit indicative of the over-all efficiency of a given connection. To meet this desirable condition, it was decided that all factors that degrade transmission be subjectively quantified in terms of equivalent decibel transmission loss.[13]

Consider two identical message channels serving the same talker and listener. On one of these introduce x dbrn* of noise. The effect will be to degrade this channel relative to the channel with no noise. Now, if one slowly degrades the "no noise" channel by introducing flat loss, there will be some value of loss for which the two circuits are conceptually equal in degradation. The actual value, say y db, is said to be the *equated transmission loss* for the given noise level of x dbrn.

The original subjective technique for the determination of these equivalent losses for noise was the articulation test. In such a test the source was a talker who uttered various selected sentences over a variable-loss test circuit to a sample of telephone users (observers) in the presence of an average amount of room noise. In each sentence there was a meaningless syllable made up of three letter sounds. For different values of circuit loss, each observer recorded the letter sounds that he thought were present in the meaningless syllables. This process was then repeated at the nominal (no loss) volume, but in the presence of various amounts of circuit noise. Next, the percentage of correctly received letter sounds (percentage of articulation) was computed for all observers, at each value of circuit loss and at each noise level. These percentages were then plotted as articulation versus loss (no noise) and articulation versus noise level (no loss). Data indicative of this procedure using 500-type sets are shown in Figs. 11(a) and 11(b), where actual values of per cent are estimated in light of the results of the original work.[6] From such curves the equated transmission losses are obtained by finding at each noise level the corresponding circuit loss that gives the same percentage of articulation. For example, 40 dbrn (noise) $\approx$ 9.0 db (loss), as is seen by comparing the 86.5 per cent articulation score.

* From here on, all units of noise measurement are in dbrn C-message.

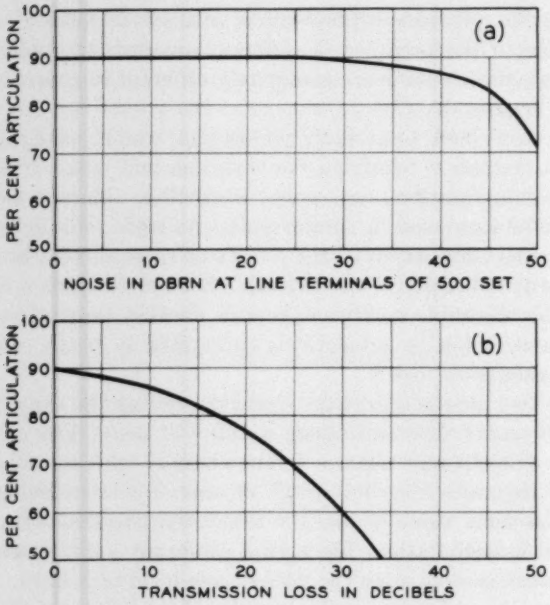


Fig. 11 — Estimated percentage of word articulation vs. (a) noise at the subscriber's telephone and (b) transmission loss.

Whether one uses articulation tests or the direct method described previously, each of the resulting equated transmission losses turns out to be invariant over a large range of received speech volume. This, of course, is an *a priori* requirement. In principle, the equated transmission loss for a given degrading effect must be independent of all others, if the sum of equated transmission losses for all effects is to be a meaningful figure of merit.

In assessing message circuit efficiency, it is useful to view equated transmission losses as transmission impairments. Consider the estimated per cent articulation data of Fig. 11. We showed, for example, that 40 dbrn of noise at the subscriber's telephone is equivalent to an approximate 9.0-db increase in transmission loss. Since this equivalence implies that 40 dbrn degrades transmission by 9.0 db, relative to transmission at maximum per cent articulation (i.e., articulation at low noise), the 9.0-db value actually constitutes the amount of *noise transmission impairment* (NTI) due to 40 dbrn of noise at the telephone. On this basis, the equated transmission losses derivable from Fig. 11 are plotted as

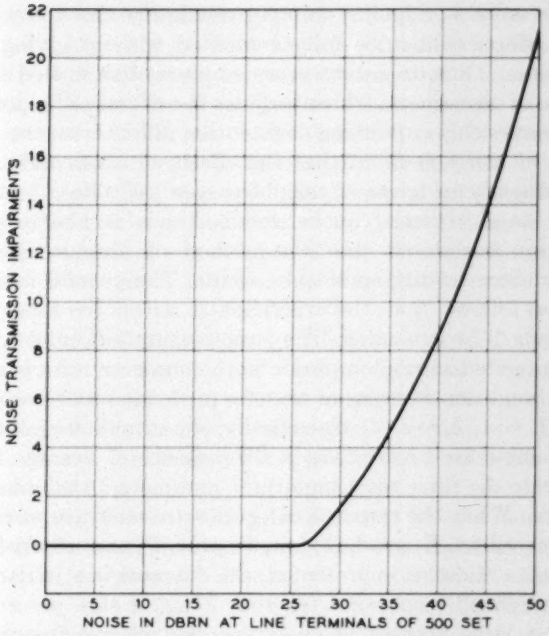


Fig. 12 — Estimated values of NTI from data of Fig. 11 under the zero NTI condition assumed in the text.

NTI versus noise in Fig. 12. For these data, a circuit with $\leqq 26$ dbrn of noise at the line terminals would be said to suffer zero NTI; a circuit with a noise level of 30 dbrn would suffer an impairment of 1.5 db.

It is quite evident, however, that NTI by themselves will not indicate noise degradation on an absolute scale. For example, if the speech-to-noise ratio on a given circuit is large, say 35 db (in equivalent units), a 5-db NTI still leaves a good effective speech-to-noise ratio, (30 db), and the transmission is still good. On the other hand, if the speech-to-noise ratio is low, say 5 db, then a 5-db NTI will be nearly fatal to transmission. This does not mean that impairments are an inefficient measure, for the same is true of actual decibels. Increasing the transmission loss of two randomly chosen circuits by 5 db does not imply that they will be equal in performance, yet they have been degraded equally. Evaluation (of this type) in decibels, including effective decibels such as impairments, always requires consideration of the "initial conditions" of the circuit under study.

From the noise standpoint, we may summarize the above by saying that noise effects cannot be fully evaluated without giving regard to speech volume. Thus, in recent years emphasis has shifted toward the second type of assessment, which includes the effects of volume directly. Here the relationship expressing degradation differs from the db-impairment type of assessment, in that the effects of noise and volume are expressed directly in terms of telephone-user attitudes.

Expressions of attitude can be obtained in a number of ways. Experimentation has shown that the method of absolute judgment[14,15] tends to produce readily applicable results. The general nature of this method is as follows: first, the experimenter selects the range and levels of the stimuli to be evaluated. The various stimuli $S_\alpha(\alpha = 1, 2, \cdots, \nu)$ are then presented in random order to the observer, who is required to judge each condition in terms of one of a predefined set of response categories R_k $(k = 1, 2, \cdots, l)$. Specifically, the stimuli are joint "received speech volume-noise" conditions in the presence of average room noise to incorporate the three most important parameters that tend to affect transmission. While the response categories are arbitrary, it is the practice to use excellent, E; good, G; fair, F; poor, P; and unsatisfactory, U.

Each time a stimulus is presented, the outcome is a judgment in one of the l predefined response categories. Thus, if this procedure is repeated for a large number of observers, the relative frequency of occurrence of a category R_k, given S_α, will tend to approach the conditional probability $p(R_k \mid S_\alpha)$; i.e., the placement of a stimulus in category k given that the stimulus is at the level α. Note, since one and only one category is assigned by every observer to each S_α,

$$\sum_{k=1}^{l} p(R_k \mid S_\alpha) = 1.$$

In contrast to equated transmission losses for noise, the results in the above procedure are probabilities of the occurrence of certain expressions of attitude. Earlier, for example, we assumed an equated transmission loss of 1.5 db for 30 dbrn at the input of the 500 set. The 1.5-db value was then adjusted to provide a measure of noise degradation in terms of NTI. Under the method of absolute judgment, we would obtain the numerical values of $p(R_k \mid S_\alpha)$ for each level of received speech volume with noise at 30 dbrn. For y vu, one might find a value of 0.2 in the E category, 0.5 in the G category, 0.2 in the F category and 0.1 in the P category. In terms of the proportion of observers responding, these numbers tend to indicate an expectancy of subscriber satisfaction — a feature not reflected directly in impairments. For example, suppose we

conduct a joint noise level-received speech volume survey in a particular community, say in locale A. Then for the joint distribution we could use either a table of noise and volume-loss impairments to determine the proportion of calls that are x db "poorer" than the reference-zero impaired call, or we could use the results of a joint noise-volume absolute judgment test to determine the proportion of calls that will be considered E, G, F, P and U. While the use of impairments gives the proportion of calls with specified values of over-all degradation, there is no direct measure of the over-all grade of transmission. Use of categorical judgments, on the other hand, provides some insight into this important question.

To illustrate the use of absolute judgments as implied above, let us assume that we have completed an absolute appraisal of joint volume-noise conditions $S_\alpha = (v_i, n_j)$, where α ranges over all combinations of volume v_i at level $i = 1, 2, \cdots, n$ and noise n_j at levels $j = 1, 2, \cdots, m$; and that S_α covers the range of any distribution of calls $p(S_\alpha) = p(v_i, n_j)$ in locale A.

Consider now any subscriber in A. In terms of the absolute judgment test, the event R_k for this subscriber is the placement of a randomly presented call in category k.* Due to the nature of R_k this event is dependent on the occurrence of some S_α. Since any S_α can occur, the event R_k occurs either as the result of both R_k and S_{11}, or both R_k and S_{12}, $\cdots$, or R_k and S_{nm}.

Symbolically, this means that

$$R_k = R_kS_{11} + R_kS_{12} + \cdots + R_kS_{nm}$$

and, since the joint events R_kS_α are mutually exclusive, their probabilities add. Thus,

$$\begin{aligned} p(R_k) &= p(R_kS_{11}) + p(R_kS_{12}) + \cdots + p(R_kS_{nm}) \\ &= p(R_k \mid S_{11})p(S_{11}) + p(R_k \mid S_{12})p(S_{12}) + \cdots \\ &\quad + p(R_k \mid S_{nm})p(S_{nm}) \\ &= \sum_\alpha p(R_k \mid S_\alpha)p(S_\alpha), \end{aligned}$$

where $p(R_k)$ is the probability that a random call S_α of joint volume-noise (v_i, n_i) is placed in category k.

If $p(R_k)$ is interpreted in the frequency sense, it gives the proportion of calls placed in category k for the given set S_α. In these terms, $p(R_k)$ is called the "grade of service" [16] for the distribution $p(S_\alpha)$.

* It is assumed here than an S_α *selected* in the experiment will be judged in the same category as the S_α appearing *randomly* at the subscribers telephone. More generally, the assumption implies that $p(R_k \mid S_\alpha)$ is stable.

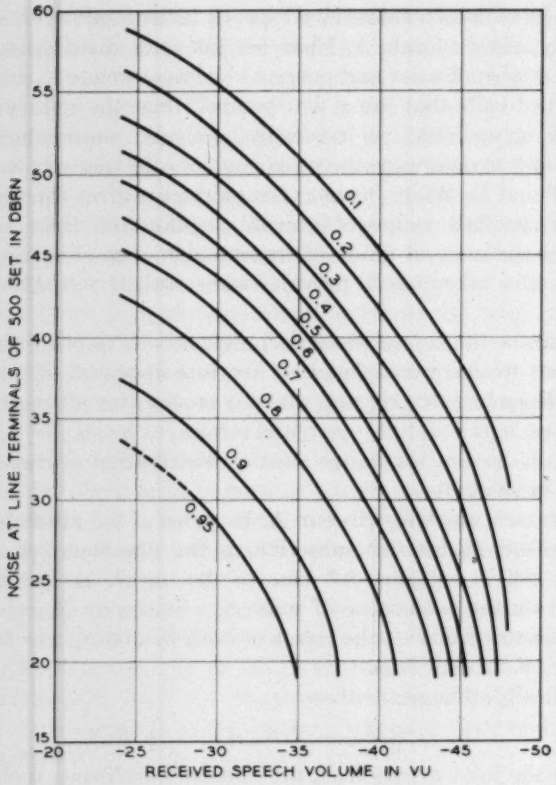


Fig. 13 — Absolute judgement: proportions of observers responding in the "good or better" category for the indicated joint volume-noise conditions.

In the more general case, where the joint volume-noise conditions S are distributed continuously with density $p(S)$, and when it is possible to obtain a continuous function $p(R_k \mid S)$ to represent $p(R_k \mid S_\alpha)$ for all α, grade of service is given by

$$\int_S p(R_k \mid S)p(S)\, dS,$$

where $S = (v,n)$.

As an example, Fig. 13 shows a part of the results of a joint volume-noise appraisal test: namely, the sum of E + G responses. Here, the conditional probabilities are of the form

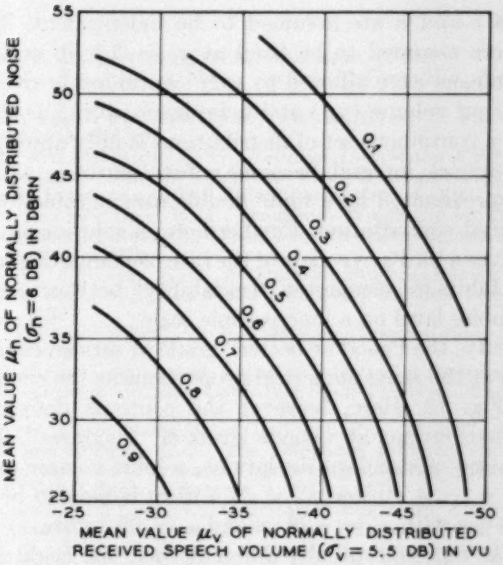


Fig. 14 — Grade of service estimates: proportions of "connections" in a bivariate normal noise-volume distribution (σ_n = 6 db, σ_v = 5.5 db) assignable to the "good or better" category when the means μ_n and μ_v have the indicated values.

$$p\left(\sum_{k=1}^{2} R_k \mid v,n\right) = p(\mathrm{E} + \mathrm{G} \mid v,n),$$

giving the probability that a volume-noise condition (v,n) is judged either excellent or good, i.e., "good or better". While such tests must be conducted at discrete levels, the incremental changes in this case were small enough to justify smoothing the data into a continuous form so that any response could be read. For simplicity, we have shown only a limited number of conditional probabilities. Notice that they are approximated by a set of contours that define equal proportions of "good or better" response in the speech volume-noise level plane. By fitting a polynomial $w(v,n)$ to these contours, the expression for "good or better" volume-noise grade of service becomes

$$\int_v \int_n w(v,n)p(v,n)\, dv\, dn.$$

An illustration of the applicability of the above is presented in Fig. 14. For $p(v,n)$ we have assumed an ensemble of bivariate normal distribu-

tions wherein v and n are assumed to be independent. The standard deviations were assumed to be fixed at $\sigma_v = 5.5$ db and $\sigma_n = 6$ db, whereas the means were allowed to vary continuously over the limited range of received volume (vu) and noise level (dbrn) as shown.*

The use of a continuous set of distributions is fully appropriate. While the grade-of-service integral becomes a function of μ_v and μ_n, it may readily be approximated by a finite double sum suitable for solution by means of digital computation. Furthermore, a solution in terms of the two means gives a bird's-eye view of the rate of change of grade of service, and it also exhibits the element of "tradability" between received speech volume and noise level on a macroscopic scale.

For simplicity, the "good or better" grade of service estimates in Fig. 14 are shown at the same numerical proportions as the conditional probabilities of Fig. 13. Here, however the contours define subjectively equivalent distributions at various levels of "goodness". Thus, for the particular choice of standard deviations, a joint volume noise distribution with $\mu_v = -34$ vu and $\mu_n = 27.4$ dbrn is seen to be subjectively equivalent to one with $\mu_v = -28$ vu and $\mu_n = 35.3$ dbrn.

In a similar but conceptually different way, one could also find subjectively equivalent volume-noise distributions from the point of view of impairments. The two cited above would not necessarily be equivalent, but, if we assume that they are, the outcome would be a table relating various proportions of these distributions to their respective impairments. As was said earlier, however, these proportions do not convey the same feeling of customer reaction as do the grade-of-service proportions. For the two joint distributions under discussion, one can expect that 80 per cent of the calls will be "good or better" — a language that has considerable appeal.

VI. ACKNOWLEDGMENTS

The authors wish to acknowledge the collective contributions over the past 50 years of the many pioneers in the field of message circuit noise evaluation. Those whose work is most familiar to us are R. S. Alford, J. M. Barstow, P. W. Blye, H. N. Christopher, D. I. Cone, O. H. Coolidge, J. W. Emling, H. E. Kent (Edison Electric Institute), H. Kahl, E. B. Mechling, H. S. Osborne and R. S. Tucker. We also wish to thank M. D. Fagen for his help, guidance and encouragement in the preparation of this article.

* Actually the integration is defined only over the surface $w(v,n)$. Thus, there is some error incurred for distribution with low and high values of mean μ_n, μ_v near the periphery of the surface.

REFERENCES

1. Cochran, W. T. and Lewinski, D. A., this issue, p. 911.
2. Harris, C. M., *Handbook of Noise Control*, McGraw-Hill, New York, 1957, Ch. 10.
3. Fletcher, H., *Speech and Hearing in Communication*, D. Van Nostrand Co., New York, 1953, Ch. 10.
4. Kent, H. E. and McCurdy, R. G., Relative Interfering Effects of Different Single-Frequency Noises in Telephone Circuits, Engineering Reports, Edison Electric Institute and Bell Telephone System, **4**, 1937, p. 163.
5. Kent, H. E. and Shetzline, R. A., Frequency Weighting for Message Circuit Noise, Engineering Reports, Edison Electric Institute and Bell Telephone System, **5**, 1943, p. 183.
6. Kent, H. E. and McCurdy, R. G., Articulation Studies on the Effects of Noise, Engineering Reports, Edison Electric Institute and Bell Telephone System, **3**, 1937, p. 255.
7. Puchstein, A. F. and Wright, C. A., *Telephone Communiation*, McGraw-Hill, New York, 1925, Ch. 17.
8. California Railroad Commission, *Inductive Interference Between Electric Power and Communication Circuits*, California State Printing Office, Sacramento, Calif., 1919.
9. Barstow, J. M., A New Noise Meter, Bell. Lab. Rec., **15**, 1937, p. 252.
10. Blye, P. W. and Gorman, L. J., The Telephone Influence Factor of Supply System Voltages and Currents, Engineering Reports, Edison Electric Institute and Bell Telephone System, **4**, 1937, p. 177.
11. Westinghouse Electric Corp., Central Station Engineers, *Electrical Transmission and Distribution Reference Book*, East Pittsburgh, Pa., 1950, Ch. 23.
12. Barstow, J. M., Blye, P. W. and Kent, H. E., Measurement of Telephone Noise and Power Wave Shape, Trans. A.I.E.E., **54**, 1935, p. 1307.
13. Emling, J. W. and McKown, F. W., A System of Effective Transmission Data for Rating Telephone Circuits, B.S.T.J., **12**, 1933, p. 331.
14. Wever, E. G. and Zener, K. E., The Method of Absolute Judgment in Psychophysics, Psych. Rev., **35**, 1925, p. 466.
15. Guilford, J. P., *Psychometric Methods*, 2nd ed., McGraw-Hill, New York, 1954, Ch. 6.
16. Coolidge, O. H. and Reier, G. C., An Appraisal of Received Telephone Speech Volume, B.S.T.J., **38**, 1959, p. 877.

A New Measuring Set for Message Circuit Noise

By W. T. COCHRAN and D. A. LEWINSKI

(Manuscript received April 15, 1960)

A new message circuit noise measuring set is described. The instrument is an easily portable, visual-indicating and direct-reading device with which rapid and accurate noise measurements may be made. The set is fully transistorized and battery powered, small in size and light in weight. It can be used for bridging, terminating and noise-to-ground measurements with a variety of plug-in weighting networks, one of which is the new C-message weighting designed for the 500-type telephone set. The sensitivity of the set is such that it can detect −90 dbm. This value is chosen as the new reference for measurement and is defined as "zero dbrn."

I. INTRODUCTION

Various forms of electrical disturbances in a telephone circuit result in acoustic noise at the output of the receiver of the telephone instrument. As noise, capable of detection by the human ear, it may be merely slightly annoying or it may be obtrusive enough to seriously impair the reception of a telephone message. If good service to the telephone user is to be maintained, it is evident that the plant forces must be provided with means for measuring electrical noise on a scale which may be correlated with transmission degradation.

An ideal instrument for this purpose would measure and combine, to a high degree of precision, those physical attributes of a noise voltage that contribute to the over-all subjective effect on the listener at the circuit terminal. Such a device would be complex indeed—so complex, in fact, that it would not be practical from an operational standpoint. However, it has been found feasible to provide a noise meter, based upon only the most important qualities of the human hearing mechanism, which has proved most useful to the plant maintenance forces. For the past 20 years the 2B Noise Measuring Set has served this purpose.

In recent years it became evident that there was a need for an im-

proved set that was simpler to use and more nearly met the needs of the present plant, and by utilizing today's advanced technology, was lighter, more portable and lower in cost. The 2B is essentially a modified version of its predecessor, the 2A Set. Each included features to permit volume and sound level measurements, which today are found in other standard instruments and are thus not needed in a noise measuring device. Furthermore, the modifications necessary to create the 2B Set from the earlier 2A added greatly to its complexity and required increased knowledge on the part of the user in order for him to obtain meaningful data. Lastly, the 2B Set is not sufficiently sensitive for many of today's uses: it does not provide adequately for the measurement of noise with the present 500-type telephone set; it is too heavy to be readily portable; and some of its components are now obsolete.

Accordingly, work was recently undertaken to develop a new noise measuring set to replace the 2B. This set, the 3A, is now in production by the Western Electric Company, and many units are already in use by the operating telephone companies. This paper will review the fundamental objectives in message circuit noise measurement, show the connection between the objectives and the requirements, and present some of the more important engineering aspects of the measuring instrument.

II. OBJECTIVES IN MESSAGE CIRCUIT NOISE MEASUREMENT*

The objective in message circuit noise measurement is to quantitatively characterize the effects of noise on the listener such that two noises that are judged to be equally interfering are assigned the same numerical magnitude. The realization of this objective in a measuring instrument imposes two types of requirements:

(a) those that imitate the important qualities of the hearing mechanism (subjective), and

(b) those that make the set compatible with the telephone system environment for the purpose of measurement (objective).

The subjective requirements must account for the manner in which noise interferes during the presence and absence of speech. This necessitates a characterization of the relative interfering effect of single-frequency noise components in the telephone channel, and an approximation of the manner in which the ear adds these components as a function of frequency and time to indicate the total effect.

Since noise effects are complicated, depending on both the character

* For a more detailed account of this topic the reader is referred to the companion paper.[1]

and type of noise, it is necessary to compromise these requirements. For example, the total interfering effect differs not only as a function of absolute noise level, but also as a function of the immediate environment (such as the room noise in the location of the telephone) and as a function of the telephone conversation level itself. Thus, it is the practice to characterize the effects of noise "on the average," and specify the requirements for the set accordingly. The characterization of relative interfering effect specifies the frequency response; the particular (average) frequency response is achieved by a weighting network. Knowledge of the pertinent integrating characteristics of the hearing mechansim determines the type of detector and the dynamic response of the indicating meter. Hence, a noise measurement made with such a device does not reflect the degree by which a given amount of noise interferes with speech transmission. Rather, the actual magnitudes are numerical entities normalized with respect to nominal conditions of noise interference.

This is not a serious drawback. After a noise set is developed and physically realized, one need only relate the measured noise magnitude to a subjective scale indicative of the degree of noise interference. The subjective tests needed for this purpose differ from the type that are used to determine relative interfering effect in that they establish the severity of measured noise level relative to ease of communication.

The objective requirements for the noise set are dictated principally by noise measuring needs. In particular, these requirements serve to transform the subjective reactions of the user (evaluated at the output of the system) to particular points in a circuit where it is most appropriate to make noise measurements. To illustrate, suppose we are in the process of deriving a weighting characteristic, that is, a subjective evaluation of the relative interfering effect of single-frequency tones as heard in a telephone. Such a test may prescribe that the observer adjust the level of each tone until it has the same interfering effect as a standard reference tone. For the measurement of noise at the line terminals of the station set, it is convenient to express the level of each of the tones in terms of relative noise voltage developed across a nominal resistance that is substituted for the telephone instrument. The relation of this voltage to frequency is the inverse of the characteristic of the desired weighting network, referred to the point of measurement (in this case, the line terminals).

The functions to be performed by a noise measuring set may be represented schematically as shown in Fig. 1. The first block consists of various inputs. As discussed above, these are chosen in conjunction with particular weighting characteristics so that the combination (input and

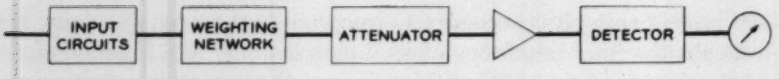


Fig. 1 — Block schematic of electronic noise measuring set.

weighting network) refer the subjective effect of noise to a convenient and appropriate point in the circuit for measurement. An attenuator and an amplifier are followed by a suitable detector, whose transfer characteristics are chosen so that the output is approximately proportional to the way the ear integrates single frequency noise components. Finally, the output is indicated on a suitable decibel meter with response characteristics similar to the ear.

III. REQUIREMENTS FOR THE 3A NOISE MEASURING SET

3.1 *General*

In accord with the fundamental objective of message circuit noise measurement, a noise measuring set must give approximately the same reading on noises that are judged to be equally interfering. In addition to this need, there are requirements that it be easy and understandable to operate.

Since the 2B Noise Measuring Set gave satisfactory answers under the conditions for which it was designed, the requirements for the 3A Set followed those of the 2B Set in principle. In addition, some of the operational features of the 2B were carried over to the 3A as well. In setting down the broad requirements, it was suggested that the new set be an easily portable, visual-indicating and direct-reading device with which rapid and sufficiently accurate noise measurements might be made.

To conform with this principle and yet adhere to the highlighting features of the 2B Set, it was decided that the new set should basically be capable of weighted and unweighted noise-metallic (bridging and terminating) and noise-to-ground measurements. In comparison to the 2B, this meant deleting volume, sound level and noise-receiver measurements. Moreover, it meant that a new weighting characteristic would have to be provided for noise measurements relative to the 500-type station set, and also that the root-sum-square law of addition (rms detection) for weighted single frequency components used in the 2B be re-examined. In addition, it was felt that the new set should be built around a flat 30-to-15,000-cycle amplifier with a variety of interchangeable plug-in weighting networks. Lastly, it was decided that the meter-

ing circuit should permit the standard 200-millisecond noise measurement and also a long-time (averaged) indication for rapidly fluctuating noise—both referred to a 1000-cycle noise reference of −90 dbm at the input.

With these considerations in mind, it was felt the new set would be sufficiently versatile to meet most of the present and anticipated future voice-frequency noise measuring needs.

3.2 *Subjective Requirements*

As discussed above, three important qualities of the hearing mechanism must be simulated in a noise measuring device: the way the ear weights single frequency noise components from the standpoint of interfering effect; the way the ear combines these frequencies to indicate the total effect; and the way the ear responds to sounds as a function of time. The first and second of these three phenomena are evaluated for average noise conditions.

Since the relative importance of single frequencies is dependent on the telephone instrument, a new weighting characteristic was derived using 500-type sets. Two tests were conducted, one in the absence and one in the presence of speech (the latter being taken at the average received level). In the absence of speech a group of observers was asked to adjust the loudness of 14 different frequencies between 180 and 3500 cycles until the sound of each was judged to be equal in annoyance to a −59 dbm 1000-cycle reference tone. Then, in the presence of speech, the same group of observers was asked to adjust the level of each of these frequencies so that the resulting impairment tended to be equivalent to the effect produced by the 1000-cycle reference tone. Throughout these experiments the room noise level was kept at 50 db referred to 0.0002 microbar.

In both cases the various single frequency levels were measured into a 900-ohm resistance, which was substituted for the 500 set. The results of each of these tests were averaged at each frequency, combined and smoothed as shown in Fig. 2.

Another test was then performed to determine the way the ear adds a larger number of single frequency noise components, weighted in accord with this particular characteristic. For earlier weighting, root-sum-square, or power addition (rms detection), was found to be a good approximation—thus the square-law detector in the 2B Set. In the present case, this form of summation was also found to be appropriate. The tests were conducted as follows:

Different pairs of bands of thermal noise were presented to various observers, who were then asked to change the level of one until it was

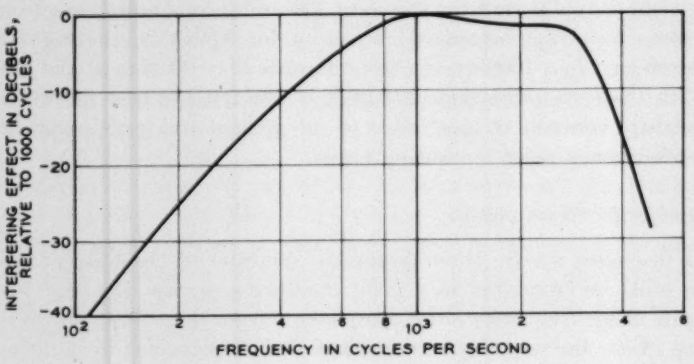


Fig. 2 — Response vs. frequency of new C-message weighting characteristic.

equal in interference to the other. These same bands were also compared objectively by computing the difference between their weighted (with the new characteristic) spectra on a power ($10^{db/10}$), voltage ($10^{db/20}$) and high-level loudness ($10^{db/40}$) basis. The best agreements between subjective and objective difference were those based on power addition. As a consequence, both power addition and the weighting shape of Fig. 2 were specified for the new set. Moreover, the same weighting was proposed for noise measurements on toll trunks in conjunction with a 600-ohm input impedance. In this respect, it was argued that, for the given weighting, the effect of distortion of the subscriber's loop would not be significant for noise originating in the toll portion of the circuit.

The third quality of hearing to be specified is the time constant of the meter circuit. Since the 200-millisecond integration time is well established,[2] this value was specified for the new set without review.

3.3 *Objective Requirements*

In general, the objective requirements for the new set were dictated by past experience with the 2B. Most important of these are (a) size and weight, (b) sensitivity and internal noise, (c) noise reference level, (d) amplifier gain and stability, (e) means of internal calibration, (f) other standard weightings, (g) inputs, (h) meter scale and attenuator steps and (i) output arrangements. For purposes of completeness, we will now discuss the particular choice of requirements for each of these features:

3.3.1 *Size and Weight*

An important consideration in the measurement of noise is the noise survey. This requires that the noise set be carried by hand for considerable distances. Measurements are made not only in telephone offices but also at cable terminals and along rights-of-way. Therefore it is most important that the noise set be small in size and light in weight. This requirement was realized in the new set by the use of miniature components and transistors. Its weight was kept to 15 pounds, its size with cover is $8 \times 11 \times 7$ inches.

3.3.2 *Sensitivity and Internal Noise*

Experience has indicated that the 2B Noise Measuring Set is not sufficiently sensitive for noise measurement on many of today's low-noise circuits. In order to assure that the new set would not have this difficulty, it was decided that the sensitivity (unweighted) be such that it could detect -90 dbm (24.5 microvolts across 600 ohms) at a maximum temperature of 50°C. In order that the new set might detect -90 dbm with some degree of meaning, the set noise (internal) is about 20 db below this figure (or -110 dbm).

3.3.3 *Noise Reference Level*

The original reference level for noise measurement with electronic measuring sets was 10^{-12} watts of 1000-cycle power at the input (-90 dbm). This reference was pertinent to noise measurements with 144-type weighting. The unit was called "dbrn" (decibels above reference noise, weighted either with 144-line or 144-receiver weighting).

With the advent of the 302-type telephone set, the 144 weightings were no longer applicable. New weightings were, therefore, derived: the F1A-line weighting for noise measurement on toll circuits, the HA1-receiver weighting for the measurement of noise across the receiver of the 302-type set. Both, however, admitted a substantially wider band of frequencies. The difference in shape was such that with the 10^{-12}-watt noise reference, the newer weightings would assign 5 db more magnitude to 3 kc flat thermal noise than the 144 weightings.

To have comparable magnitudes (for approximately equal interfering effect), a new reference was proposed for the newer weightings. This reference was a 5 db upward adjustment from 10^{-12} to 3.16×10^{-12} watts. The new reference was designated dba (dbrn adjusted).

Since the C-message weighting admits a wider equivalent spectrum than does the F1A line weighting, a similar approach would require a

still higher (less sensitive) equivalent reference. This was considered for the new set, but it was deemed unwise. Using a higher equivalent reference with a more sensitive device would result in "negative" noise magnitudes on low noise circuits. To circumvent this undesirable outcome, it was decided to revert to the lower 10^{-12}-watt 1000-cycle reference for the new 3A Set. While this reference is numerically equal to the original one at 1000 cycles, and while the scale in the new set is designated dbrn, the two dbrn units are not equivalent when viewed in terms of weighted noise measurement. It is important, therefore, to designate all C-message weighted measurements with the new set as "dbrn C-message".

3.3.4 *Amplifier Gain and Stability*

For the given sensitivity it was estimated that the net voltage gain of the amplifier should be approximately 90 db. This was to assure adequate output signals for metering, monitoring and recording purposes on any measurable input signal. At full scale deflection, the open-circuit ac output is approximately 1.2 volts rms.

While the amplifier in the 2B Noise Measuring Set has fairly good long-time stability, it is usually necessary to recalibrate about every hour. In the new set, increased amplifier stability is realized by improved circuit techniques. It is not necessary to calibrate the instrument more than twice each day.

3.3.5 *Means of Internal Calibration*

The new 3A (as does the 2B Set) provides for internal calibration at 1000 cps. The reason for this choice of frequency is twofold: (a) the insertion loss for all frequency weightings is the same at 1000 cps, and (b) this is the frequency of the standard milliwatt reference generator, found in most offices, which provides the means for primary calibration of the set.

3.3.6 *Weighting Characteristics*

In addition to the basic C-message weighting, it is desirable to make various other weighted noise measurements. For flexibility, these are packaged to be used on a "plug-in" basis. Past experience has indicated the need for 3-kc flat, 15-kc flat and 5-kc weighted program noise measurements. These should provide for all message circuit noise measuring needs.

The nominal frequency response of each of these weightings is given in Figs. 2, 3 and 4.

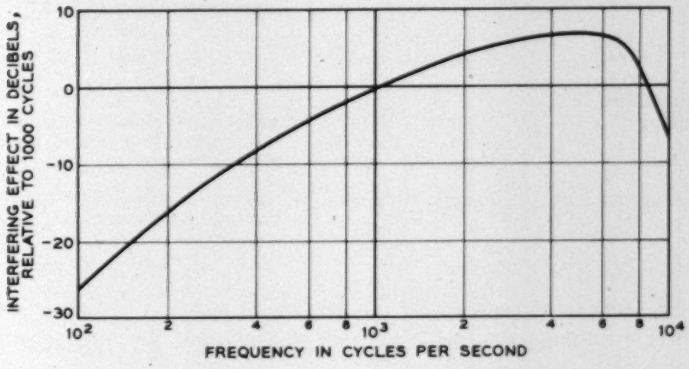


Fig. 3 — Response vs. frequency of standard 5-kc program circuit weighting.

3.3.7 *Inputs to the New Set*

Circuit arrangements for the inputs to the set are similar to many found in the 2B. The inputs are listed in Table I, in conjunction with pertinent weightings (and their purpose).

The longitudinal input circuit balance for each of the metallic input arrangements is more than 85 db at 60 cycles and more than 55 db at 15 kc.

Each input is obtained by connecting the circuit under test through a

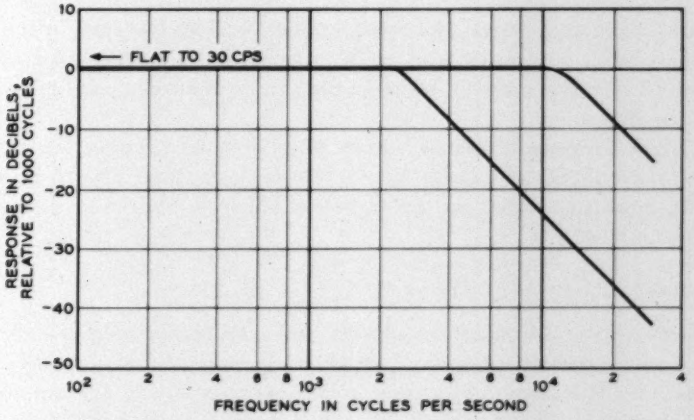


Fig. 4—Response vs. frequency of 3-kc flat and 15-kc flat weighting characteristics.

TABLE I

Input	Input Impedance, ohms	Purpose
Noise-metallic (N_m)		
Bridging	Nominal 10,000	Weighted or unweighted noise-metallic measurements on message or program circuits
600	Nominal 600	
900	Nominal 900	
Noise-to-ground (N_g)	Nominal 100,000	Weighted or unweighted noise-to-ground measurements on open wire or cable facilities

standard telephone jack or by connection to a pair of binding posts, selecting the input by a function switch and inserting the desired weighting network.

3.3.8 *Meter Scale and Attenuator Steps*

From an operational standpoint the meter scale-attenuator feature in the new set is like that in the 2B. The major difference is that the meter scale is calibrated over a zero-to-12 db range and the attenuator in 5 db steps from 0 to 85 db. The 12-db range of the meter resulted from calibrating as much of the usable range of the meter as practicable. This arrangement allows noise measurements from 0 to 97 dbrn. As mentioned earlier, no corrections are needed for any one of the various possible measurements.

3.3.9 *Output Arrangements*

As we have seen above, the measurement of noise has qualitative aspects. It is therefore desirable to be able to listen to the noise being measured. For this purpose, an ac output and a headphone are provided with the set.

It is also convenient in noise survey work to make a continuous record of the noise levels at a point in the telephone system. Therefore a dc output is also provided on the new set, which is capable of driving a variety of recorders.

IV. CONSIDERATIONS OF TRANSISTORS

To make the 3A Noise Measuring Set a truly portable device, the use of transistors as the active elements was investigated as a first consideration. The minimum power to be measured made the transistor noise figure of primary concern. It is rather elementary to find the maximum permissible noise figure for the set—and consequently for the initial

stage of amplification—since we know the three necessary numbers: the minimum power to be measured (10^{-12} watts), the maximum bandwidth to be used (15 kc), and the maximum operating temperature (50°C).

Thermal power is equal to kTB, where k is Boltzman's constant, T is absolute temperature and B is bandwidth. Numerically, this is $(1.38 \times 10^{-23})(323)(15{,}000)$, which is approximately 6.7×10^{-17} watts or -131.7 dbm. If the measuring set is to measure -90 dbm, then the internal noise should be 20 db below this number (even the 20-db difference will slightly affect the accuracy of the set). This results in a maximum noise figure of $(-90 - 20) + 131.7 = 21.7$ db. This is certainly not an unrealistic noise figure requirement for the noise measuring set, but it is, nevertheless, low enough that it could not be neglected in the design.

Requirements other than noise figure also influence the consideration of transistors as the active elements in the set. Such things as bandwidth (15 kc), voltage gain (about 90 db), and the required temperature and time stability have been obtained earlier by many people using present-day feedback techniques. The only remaining quantity in the list of requirements is the proposed output voltage and overload margin. The requirements indicate that one volt (rms) open circuit at the output is desired for a full-scale meter indication. They also indicate that the set should not overload for signals 20 db higher than full scale. This means, then, that the output voltage should be as high as 10 volts (about 14 volts peak) without distortion. These numbers, in themselves, are by no means impossible, but it was found during the development of the set that they necessitated a very large battery. Since this would defeat the purpose of using transistors (portability), the overload requirements were eased to 15 db.

V. THE DEVELOPMENT OF THE 3A

The general noise measuring device consists of six black boxes: input circuit, attenuator, amplifier, frequency weighting, detector and meter. The development problem might be defined as designing these black boxes and physically ordering them in the measuring set. The input circuit, by its nature, is first, and the detector and meter are last. This leaves the attenuator, the amplifier and the frequency weighting yet to be placed.

We have seen that the amplifier must have a relatively low noise figure (about 21 db maximum). A low noise figure may be achieved by satisfying certain bias conditions on the first stage of amplification.

Montgomery[3] has shown that low base-to-collector voltage and low emitter current lead to a low noise figure for an alloy transistor. (Since this work was done, lower noise transistors have become available, which would have rendered this problem less severe.) These low-bias conditions impose the necessity of good dc stability. Also required of the amplifier is a high input impedance, since the new set is to be basically a bridging instrument. These factors, and the realization that the design of feedback amplifiers with more than three stages becomes somewhat burdensome, dictated the use of two amplifiers.

The first of these was built with the noise figure as the primary objective. In doing this, and striving for sufficient feedback for the necessary stability, the voltage gain was set at about 25 db. This, with a 9-db loss allotted to the frequency-weighting networks, sets the voltage gain of the second amplifier at 74 db (giving the required 90 db over-all voltage gain). The two amplifiers were built, with an eye for economy, using standard techniques to obtain the required bandwidth, stability and gain characteristics.

Because of the bias conditions on the first stage of amplification it follows that the input signal should be kept quite small in order to avoid nonlinear operation. This consideration places the attenuator between the input circuits and the first amplifier

It is not advisable to put the weighting network ahead of the first amplifier, because its loss will degrade the noise performance of the set. Furthermore, the input of the first amplifier provides a better termination for the attenuator than would the network. Finally, terminating the weighting networks in the two amplifiers eases the network design problem. The final layout, then, becomes as shown in Fig. 5.

The input circuit has associated with it a switch, controlled from the front panel, known as the function switch. The first position is OFF. In this position the battery is disconnected and the meter is shorted for protection in handling. The next two positions provide battery and calibration checks. They have been positioned here to encourage the user to make these checks when turning the set on. The remaining posi-

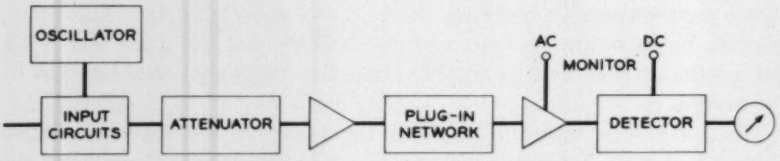


Fig. 5 — Block schematic of 3A Noise Measuring Set.

tions select one of four input conditions: N_g (noise-to-ground measurements), BRDG (bridging input impedance 10,000 ohms), 600 and 900 (terminating impedance of 600 ohms or 900 ohms).

Another switch on the set, called the DBRN switch, controls the loss in the attenuator. The attenuator is constructed of simple pi and tee networks of resistors, which are combined by the switch to give loss from 0 to 85 db in 5-db steps. Again with an eye on economy, the attenuator was built using only 14 resistors and a three-deck switch for its construction. The 5-db increment was determined by the largest practical logarithmic scale on the meter. Since this is about 10 db, the 5-db attenuator increments are necessary for convenient meter readings.

In order to give assurance that the meter reading is correct, an internal calibration source is provided. This source is a 1000-cycle transistorized Hartley-type oscillator using the common collector configuration. A breakdown diode is used to maintain the dc base-to-collector voltage at a nearly constant value. The level of oscillation is determined by this voltage, and is consequently quite stable.

The networks, which were designed to be used on a plug-in basis with the set, were developed using standard network synthesis techniques. The nominal frequency characteristics of these networks are pictured in Figs. 2, 3 and 4. These networks, of course, make the difference between the 3A Set being a voltmeter and its being a noise measuring set. As such, they form one of the two basic elements of the set, the other being the detector.

As discussed earlier, the proper summation of the spectral components of the noise can best be approximated by root-sum-square addition. A detector responding to the root-mean-square value of the complex waveform resulting from these components will accomplish this. The other restriction calls for a 99 per cent response to a noise burst in 200 milliseconds. Thermocouples and thermistors, which have rms response, are too slow for this application. The desired response was therefore approximated by a rectifier circuit. This type of detector was used in the 2B Set, but was achieved through the use of selected diodes. Analysis showed that it was possible to design a rectifier of satisfactory characteristics. This is covered in the Appendix.

The complete set is shown in Fig. 6. Its performance meets all requirements for a modern noise measuring set.

VI. CONCLUDING REMARKS

While the instrument described here performs the job of noise measurement, there are other features of importance in the design of a noise

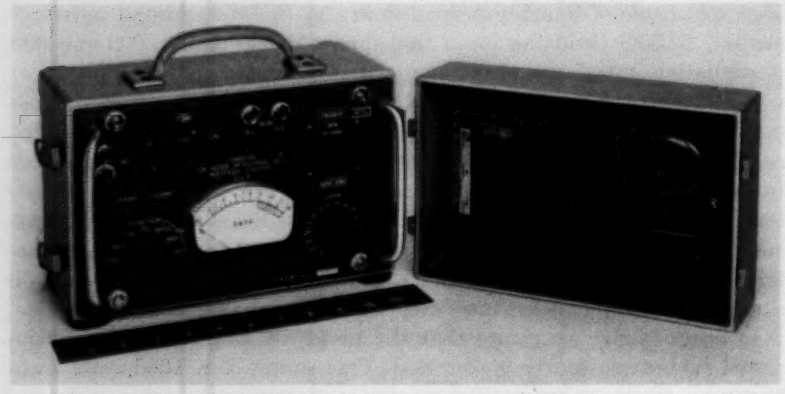


Fig. 6 — The 3A Noise Measuring Set.

set for practical application in the telephone industry. Three of these features are cost, operational simplicity and versatility. Of these, cost was possibly the most important element influencing the development of the set. But certain things cannot be overlooked because of their initial "cost." One of these things is operational simplicity. This is achieved through at least two considerations. One is keeping the number of controls down and their arrangement logical. The other is the arrangement of the operating steps in a logical order. Perhaps both of these can best be demonstrated by describing how a measurement is made.

After the operator connects the circuit to be tested to the input of the noise set, he selects the proper weighting network and plugs it into the set. He then turns the function switch through the BAT and CAL positions (assuring the set is in proper working order) to one of the four input impedances. Finally, he turns the attenuator switch (called the DBRN switch) until he can observe a meter indication. The noise reading is then the sum of the DBRN switch indication and the meter indication. A typical reading might be "23 dbrn—C-message", where 23 dbrn is the reading and C-message is the frequency weighting that was used.

TABLE II

Type of Noise	2B, dba	3A, dbrn	Difference, db
Power harmonics, 180, 300 and 540 cycles	20	25	5
Modulation product noise	20	27	7
Panel office switching noise	20	26	6
3-kc thermal noise, power hum and switching noise	20	27	7

Note, however, that the 3A Noise Measuring Set will read higher (using C-message weighting) than the 2B Noise Measuring Set (using F1A-weighting), as shown in Table II. Although the numerical difference depends on the character of the noise, it is found in general that the reading is about 6 db higher on the 3A. This relationship should be remembered during the transition period while the 3A is superseding the 2B.

The final item of interest in the design of the set is versatility, which is exemplified by the frequency-weighting networks. For instance, if a new frequency weighting should become necessary due to some change in human attitudes, the character of the interference, or the noise sensitivity of the telephone plant, a new network could be designed and built into a package that would fit into the present set. Another aspect of versatility is reflected in the provision of ac and dc outputs, which may be used to drive a variety of other instruments.

VII. ACKNOWLEDGMENTS

The authors wish to thank D. W. McLellan for performing the measurements relating the performance of the 3A to the 2B. We also wish to acknowledge the help given by A. J. Aikens and T. C. Anderson in the preparation of this paper. The networks used in the 3A Set were synthesized by H. Simon and H. M. Thomson.

APPENDIX

Suppose that $f(t)$ is a voltage applied to the input of the circuit of Fig. 7(a). If the RC product is much larger than the period of the lowest spectral component of $f(t)$, then the circuit will "peak-detect" the input voltage. For this case $\hat{E}$ will be the dc voltage corresponding to the peak value of $|f(t)|$.

If the same signal is applied to the input of Fig. 7(b), and if $\bar{E}$ is the dc voltage across R, then $\bar{E}$ corresponds to the average of $|f(t)|$. This might be called "average" detection.

Since the rms value of a positive function lies between the average and peak value, it is instructive to investigate the action of a detector which gives a dc voltage corresponding to something between average and peak. It might be termed "quasi-rms". Such a detector might be built by properly combining Figs. 7(a) and 7(b). The result is given in Fig. 7(c).

In Fig. 7(c), let $\bar{E}_1$ and $\bar{E}_2$ be the dc components of the voltages across R_1 and R_2, respectively. If the R_2C product is long (as above), then the voltage across R_2 is practically constant and equal to $\bar{E}_2$.

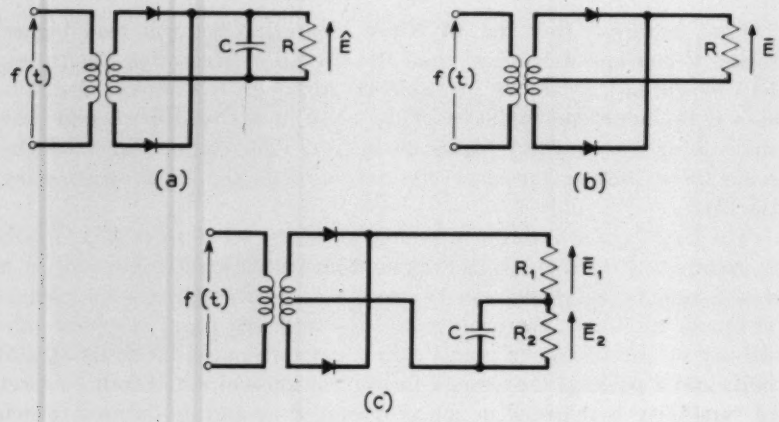


Fig. 7 — (a) Peak detection; (b) average detection; (c) "quasi-rms" detection.

Then, when $|f(t)|$ is less than $\bar{E}_2$, the diodes are back-biased and $\bar{E}_1$ is zero. The waveform across R_1 and R_2 is as pictured in Fig. 8(a) (ideally). Since $\bar{E}_2$ is the voltage across R_2, then $\bar{E}_1$ is the average of the unshaded portion of Fig. 8(a), or

$$\bar{E}_1 = \frac{1}{T}\int_{t_1}^{t_2} [|f(t)| - \bar{E}_2]\, dt, \tag{1}$$

where $f(t_1) = f(t_2) = \bar{E}_2$. Assume, for simplicity,

$$f(t) \geqq \bar{E}_2 \qquad \text{when } t_1 \leqq t \leqq t_2,$$
$$f(t) < E_2 \qquad \text{otherwise.}$$

This is not a necessary assumption, since the integral in (1) could be expressed so that the assumption is true.

Let $\bar{E}$ be the detected voltage; i.e,

$$\bar{E} = \bar{E}_1 + \bar{E}_2.$$

Then, from (1),

$$\bar{E} = \bar{E}_2 + \frac{1}{T}\int_{t_1}^{t_2} [|f(t)| - \bar{E}_2]\, dt. \tag{2}$$

Now, let E_0 be the true rms value of $f(t)$ and let $KE_0 = \bar{E}$. Then K is the factor by which $\bar{E}$, the detected voltage, differs from E_0, the true rms voltage. From (2):

$$KE_0 = \bar{E}_2 + \frac{1}{T}\int_{t_1}^{t_2} [|f(t)| - \bar{E}_2]\, dt. \tag{3}$$

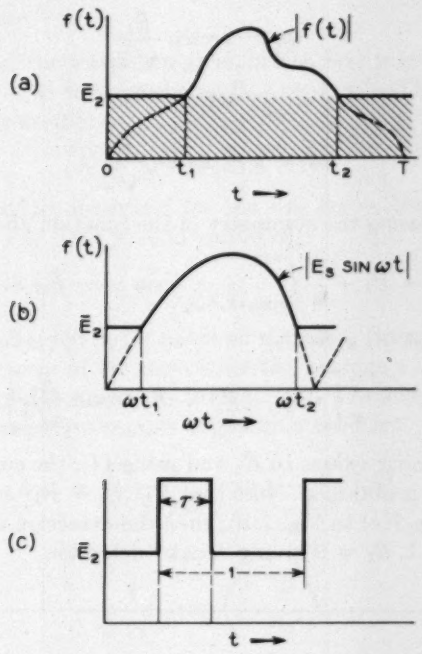


Fig. 8 — (a) Voltage across R_1 and R_2 with $f(t)$ into the detector; (b) sine wave into the detector; (c) rectangular wave into the detector.

If we let $E_0 = 1$ for all $f(t)$'s, then for a given $f(t)$ we have K in terms of $\bar{E}_2$. There is another relation between K and $\bar{E}_2$, which may be found by observing from Fig. 7(c) (and $E_0 = 1$) that

$$\frac{\bar{E}_2}{\bar{E}_1 + \bar{E}_2} = \frac{\bar{E}_2}{K} = \frac{R_2}{R_2 + R_1}. \tag{4}$$

The problem is now to find $R_2/(R_1 + R_2)$ such that $K = 1$ for a given $f(t)$. Possibly the simplest procedure is to proceed graphically by first finding K in terms of $\bar{E}_2$ from (3) and assigning values to $\bar{E}_2$ to determine K. Then, from (4), $R_2/(R_1 + R_2)$ may be found. Finally, plot K versus $R_2/(R_1 + R_2)$. Several examples follow:

Sine Wave

In Fig. 8(b), the amplitude of the sine wave, E_s, is given as $\sqrt{2}$ since we want $E_0 = 1$. From the figure, it is clear that

$$\omega t_1 = \arcsin \frac{\bar{E}_2}{\sqrt{2}}$$

and

$$\omega t_2 = \pi - \arcsin \frac{\bar{E}_2}{\sqrt{2}}.$$

Then, from (3), using the symmetry of the function about $\pi/2$, we have

$$K = \bar{E}_2 + \frac{2}{\pi} \int_{\arcsin (\bar{E}_2/\sqrt{2})}^{\pi/2} [\sin \omega t - \bar{E}_2]\, d(\omega t). \tag{5}$$

The evaluation of (5) gives:

$$K = \frac{2}{\pi} \left[\sqrt{2 - \bar{E}_2^2} + \bar{E}_2 \arcsin \frac{\bar{E}_2}{\sqrt{2}} \right]. \tag{6}$$

By simply assigning values to $\bar{E}_2$ and using (4), the curve labeled "sine wave" on Fig. 9 is obtained. Note that $R_2/(R_1 + R_2) = 0$ gives $R_2 = 0$. This reduces Fig. 7(c) to Fig. 7(b); then the detection is "average." For $R_2(R_1 + R_2) = 1$, $R_1 = 0$ giving "peak" detection.

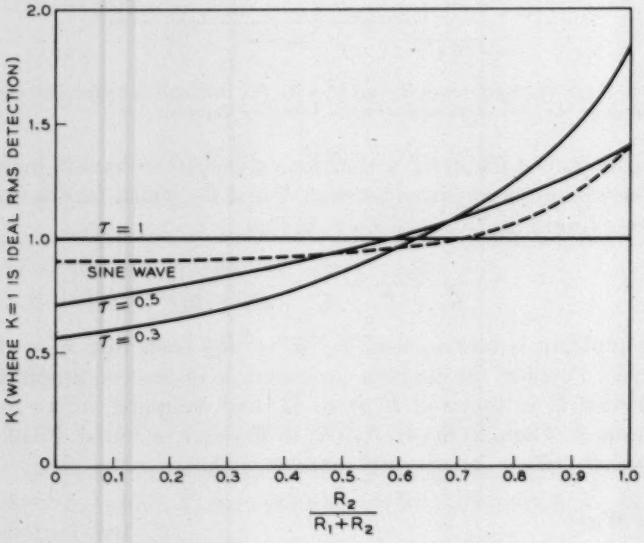


Fig. 9 — Response of the detector to a sine wave and rectangular waves of duty cycle τ.

Rectangular Wave

The rms value of the waveform given in Fig. 8(c) is given by $E_{max}\sqrt{\tau}$. Since we want $E_0 = 1$, then we choose $E_{max} = 1/\sqrt{\tau}$. The expression for K easily follows from (3):

$$K = \sqrt{\tau} + E_2(1 - \tau)$$

Using the procedure described for the sine wave, the curves labeled "τ" are given on Fig. 9 for several values of τ.

Gaussian Noise

For gaussian noise, $f(t)$ in (3) holds no meaning. Remember, however, that the integral term of (3) represents the average value, or dc component, of the voltage across R_1 in Fig. 7(c). The average value of a doubly rectified long-term sample of gaussian noise is:

$$E_{ave} = \int_0^\infty \frac{2e}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{e^2}{2\sigma^2}\right) de, \tag{7}$$

where

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{e^2}{2\sigma^2}\right)$$

is the probability density function for the voltage e. The factor of 2 is present, since double rectification simply "folds" the negative half of the density function across the mean value of e (which is zero). The rms value of such a signal is σ. Since we want E_0 (the rms value) = 1, we set $\sigma = 1$. The average value of the voltage across R_1 in Fig. 7(c) can be expressed in a fashion similar to (7). In this case, however, $\bar{E}_2$ is subtracted from the value of the instantaneous noise voltage, e. Then (7) becomes

$$\bar{E}_1 = \int_{\bar{E}_2}^\infty (e - \bar{E}_2) \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{e^2}{2}\right) de, \tag{8}$$

where the lower limit is $\bar{E}_2$, since the voltage across R_1 cannot be negative.

From (3) we have:

$$K = \bar{E}_2 + \int_{\bar{E}_2}^\infty (e - \bar{E}_2) \sqrt{\frac{2}{\pi}} \exp\left(-\frac{e^2}{2}\right) de. \tag{9}$$

Evaluation of (9) gives:

$$K = \sqrt{\frac{2}{\pi}}\left[\exp\left(\frac{\bar{E}_2^2}{2}\right) + \bar{E}_2 \int_0^{\bar{E}_2} \exp\left(-\frac{e^2}{2}\right) de\right].$$

As in the two previous examples, values were assumed for $\bar{E}_2$, and K was determined. Then, from (4), $R_2/(R_1 + R_2)$ was found using $\bar{E}_2$ and K. The results for gaussian noise are given in Fig. 10.

Conclusion

Similar curves may be plotted for a wide variety of functions using the techniques given in the examples. Fig. 11 is an illustration of the curves given for many examples including those of this Appendix. However, since the detector will be calibrated with a sine wave, the curves of Fig. 9 are referred to a sine wave. The figure represents, then, the error in the detection of the rms value of a set of functions when the detector is calibrated to read correctly the rms value of a sine wave. The error is given in decibels.

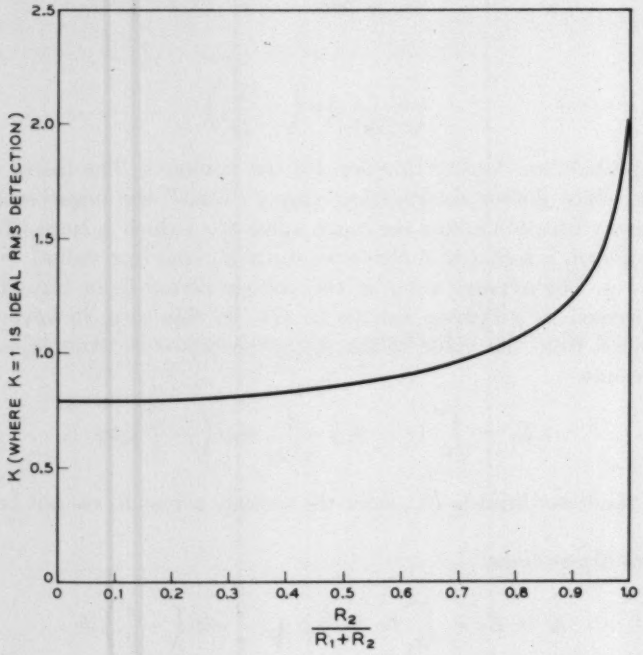


Fig. 10 — Response of the detector to gaussian-distributed noise.

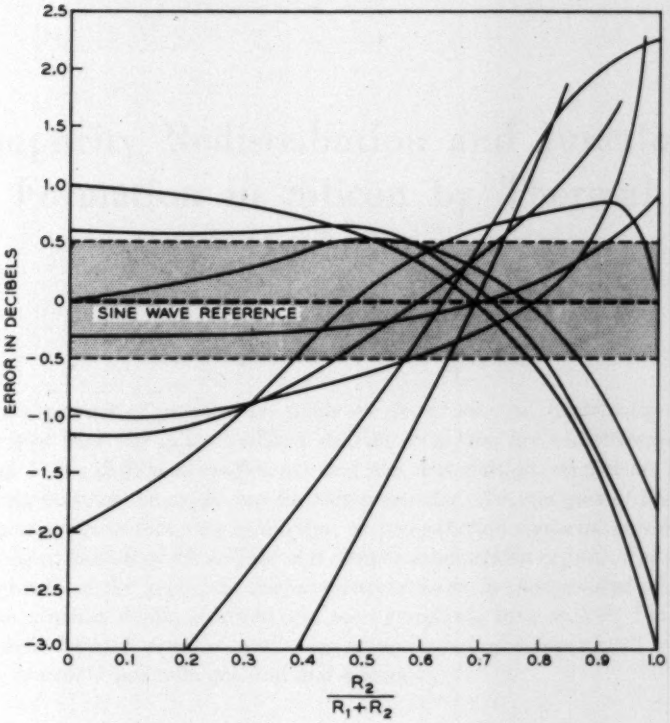


Fig. 11 — "Quasi-rms" detector error for several functions.

With only a few exceptions, the curves of Fig. 11 have a striking similarity in that they all fall within 0.5 db of "true rms" for $R_2/(R_1 + R_2)$ equal to about 0.6. Thus, if the resistors of Fig. 7(c) are properly chosen it is theoretically possible to detect the rms value of many different waveforms with an error as small as 0.5 db.

Although the assumptions of a perfect transformer, perfect diodes and $\bar{E}_2$ to be a constant may seem to be restrictive in the analysis, experimental results verified the calculations.

REFERENCES

1. Aikens, A. J. and Lewinski, D. A., this issue, p. 879.
2. Munson, W. A., The Growth of Auditory Sensation, J. Acoust. Soc. Am., **19**, 1947, p. 584.
3. Montgomery, H. C., Electrical Noise in Semiconductors, B.S.T.J., **31**, 1952, p. 950.

Impurity Redistribution and Junction Formation in Silicon by Thermal Oxidation

By M. M. ATALLA and E. TANNENBAUM

(Manuscript received January 28, 1960)

In the process of growing an oxide on doped silicon, electrically active impurities near the silicon/silicon dioxide interface are redistributed according to the diffusion coefficients and the distribution coefficient of the impurity between the oxide and the semiconductor. An analysis of the phenomenon predicts that single-junction or two-junction material can be obtained by oxidation of the surface of a compensated silicon crystal. For parabolic growth of the oxide, the surface concentration is independent of time, and the junction depth, gradient and sheet resistivity vary with $t^{\frac{1}{2}}$. This has been demonstrated experimentally by oxidation of a compensated p-type silicon crystal doped with gallium and antimony.

I. INTRODUCTION

During the growth of an oxide on a semiconductor surface, the doping impurities of the crystal may undergo a redistribution in the semiconductor in a region near the surface. The extent of the redistribution, which may range from a pile-up region to a depletion region, depends on the segregation coefficient of the impurity at the semiconductor-oxide interface, on its diffusion coefficients in both the semiconductor and the oxide and on the rate and time of oxidation.

To illustrate, consider the case where an impurity is completely rejected by the oxide; i.e., its segregation coefficient k is zero. The resulting concentration gradient will cause diffusion of the impurities back into the body of the semiconductor. This is a nonequilibrium problem, where the resulting impurity concentration profile will change continuously with time. Fig. 1 shows schematically two redistribution profiles for two extreme cases of pile-up ($k < 1$) and depletion ($k > 1$).

It is evident that, with more than one impurity in the semiconductor,

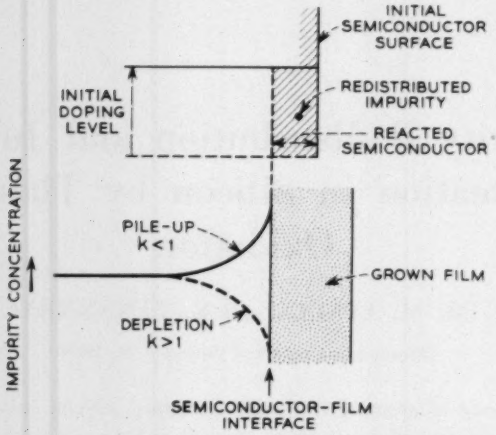


Fig. 1 — Schematic diagram of the redistribution of an impurity, originally uniformly distributed, produced by thermal oxidation of semiconductor surface.

this process may produce various distribution profiles with single or multiple junctions. Fig. 2 illustrates two such possibilities.

In this paper we will present: (a) an analysis of the impurity redistribution phenomenon and its application to junction formation, and (b) experimental data for the system silicon-silicon dioxide with gallium and antimony as impurities.

II. THEORY

In Section 2.1 the pertinent equations and boundary conditions of a model of the redistribution process is presented. In Section 2.2, an ap-

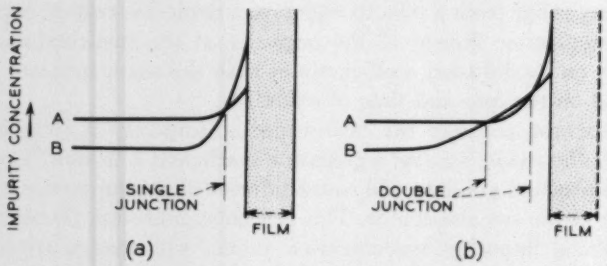


Fig. 2 — Schematic diagram of the concentration profiles of two impurities produced by thermal oxidation: (*a*) single junction formation; (b) double junction formation.

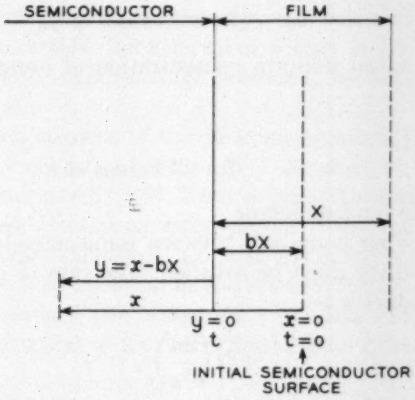


Fig. 3 — Coordinates and notations used in the analysis of the impurity redistribution phenomenon.

proximate solution is given for an important practical case where the segregation coefficient of the impurity at the semiconductor-film interface is negligibly small. In Section 2.3 an approximate solution is given for the case of parabolic growth of film thickness with time and for any segregation coefficient. In Section III, the above solutions are applied to determine some junction characteristics.

2.1 *Model of Redistribution Process*

Fig. 3 shows the coordinates and some of the terminology used below. At time $t = 0$, the semiconductor boundary corresponds to the plane $x = 0$ when the film growth process is started. At any time t let the film thickness be X. The corresponding thickness of semiconductor material used in producing this film is bX, where b is a constant. For the system silicon-silicon dioxide, b is about 0.44 for the amorphous oxide. Using the moving coordinate $y = x - bX$, where $y = 0$ corresponds to the instantaneous location of the semiconductor-film interface, the diffusion equation for any point within the semiconductor takes the form

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial y^2} + b\frac{dX}{dt}\frac{\partial n}{\partial y}, \tag{1}$$

where n is the impurity concentration and D is the diffusion coefficient of the impurity in the semiconductor. The term dX/dt is the rate of film growth that must be obtained from experimental data. The boundary conditions that must be satisfied are:

i. At $t = 0$,

$$n = n_0 \qquad \text{for all values of } y, \tag{2}$$

where n_0 is the initial uniform concentration of impurity in the semiconductor.

ii. At $y = \infty$,

$$n = n_0 \qquad \text{for all values of } t, \tag{3}$$

for a semi-infinite semiconductor.

iii. At the moving boundary between semiconductor and film, conservation of impurity must be satisfied. The rate of impurity diffusion into the semiconductor is

$$-D\left(\frac{\partial n}{\partial y}\right)_0,$$

For a segregation coefficient k, the rate at which impurities are admitted into the oxide is

$$kn_s\left(\frac{dX}{dt}\right),$$

where n_s is the instantaneous surface concentration of the impurity in the semiconductor.* For conservation of impurity, the sum of the above terms must be equal to the impurities in the reacted semiconductor $n_s[b(dX/dt)]$:

$$D\left(\frac{\partial n}{\partial y}\right)_0 + n_s(b - k)\frac{dX}{dt} = 0. \tag{4}$$

Measurements by Ligenza[3] of the kinetics of oxidation of silicon under various conditions have shown that the relation between film thickness and time can be represented by the following power expression:

$$X^m = Kt, \tag{5}$$

where K is the oxidation constant that is dependent on the pressure of the reacting gas and on temperature. The temperature dependence is exponential.

The activation energy as well as the exponent m are dependent on the

* This implicitly assumes that the film growth takes place at the semiconductor-film interface. Evidence for such a process has been shown for the oxidation of silicon under various conditions.[1,2] Furthermore, if a film grows at the film gas interface, i.e., by migration of the semiconductor atoms or ions through the film, and if the diffusion coefficient of the impurity in the film is negligibly small, one obtains the case of effectively complete rejection of the impurity by the film. If, on the other hand, the impurity diffusion in the film cannot be neglected, one must include an additional diffusion equation for the impurity in the film and solve it simultaneously with (1). Such a case was not considered in this paper.

oxidation conditions. For oxidation in oxygen or water vapor at pressures of one atmosphere or less, the activation energy is 1.7 ev and m is 2. For oxidation in water vapor at high pressures, the activation energy is 1.0 ev and m is 1.0.

Now one desires a solution of the diffusion equation, (1), that satisfies the three boundary conditions given, (2), (3) and (4), with the oxidation process as described by (5). Such a general solution was not obtained. Approximate solutions, however, were derived based on discarding the diffusion equation, (1), and assuming a time-dependent exponential impurity distribution in the semiconductor. Such a solution satisfies all the boundary conditions, (2), (3) and (4), as well as the condition of conservation of all redistributed impurities.*

2.2 *Impurity Redistribution by a Film Growth $X^m = Kt$ with a Segregation Coefficient $k = 0$.*

At any time during the process, it is assumed that the impurity distribution in the semiconductor follows the following exponential form:

$$n - n_0 = (n_s - n_0)e^{-y/A}, \tag{6}$$

where n_s and A are in general time-dependent. This equation already satisfies our second boundary condition, (3). To satisfy the third boundary condition, (4), one substitutes (6) in (4) and obtains

$$1 - \frac{n_0}{n_s} = \frac{A}{D}\left(\frac{dX}{dt}\right)b. \tag{7}$$

Furthermore, for the conservation of impurity,

$$\int_0^\infty (n - n_0)\,dy = n_0 bX.$$

Substituting in (6) and integrating, one obtains:

$$\frac{n_s}{n_0} - 1 = \left(\frac{b}{A}\right)X. \tag{8}$$

From (7) and (8) one solves for n_s and A, which are obtained as functions of time. It is evident that the first boundary condition, (2), is also satisfied:

$$\frac{n_s}{n_0} = P + (P^2 - 1)^{\frac{1}{2}} \tag{9}$$

* More recently, Doucette et al., have obtained a rigorous solution for the special case of parabolic film growth ($m = 2$). Concentration profiles based on the above approximate solution deviate by less than 10 per cent from exact profiles.

and

$$A = \frac{bX}{\frac{n_s}{n_0} - 1},$$

where

$$P = 1 + \frac{b^2}{4D}\frac{d(X^2)}{dt} = 1 + \frac{b^2}{4D}\left[\frac{2}{m}K^{2/m}t^{2/(m-1)}\right]. \tag{10}$$

From the above equations and the distribution equation, (5), one obtains the desired distribution of impurities in the semiconductor as a function of time. It is of interest to note that both the surface concentration n_s and the penetration A/X, where X is the film thickness, are determined by the single parameter P, as defined by (10). Except for a parabolic growth of the film, where $m = 2$, the parameter P is generally dependent on time; i.e., the surface concentration n_s and the penetration A/X are time-dependent. For parabolic growth, however, the surface concentration n_s is independent of time and the penetration A/X is also independent of time; i.e., the penetration of the redistribution A is directly proportional to the film thickness X. For the two practical cases of parabolic and linear growth one, therefore, obtains

$$P = 1 + \frac{b^2K}{4D} \quad \text{(parabolic growth)} \tag{11}$$

and

$$P = 1 + \frac{b^2K^2}{2D}t \quad \text{(linear growth)}. \tag{12}$$

The temperature dependence of the process appears in the ratio K/D for parabolic growth. Since both K and D are exponentially dependent on temperature, the process of impurity distribution is also exponentially dependent on temperature, with an activation energy equal to the difference between the activation energies of the impurity diffusion in the semiconductor and the film growth process.

For illustration, the above equations were used to calculate the surface concentration of redistributed impurities in silicon when it is thermally oxidized in oxygen. The oxidation in this case is parabolic ($m = 2$) and the oxidation constant K is best fitted in the temperature range of 900 to 1200°C and atmospheric pressure* by the following expression:

* The oxidation constant K, according to Ligenza,[3] is proportional to $p^{0.8}$, where p is the oxygen pressure.

$$K = 8.3 \times 10^{-8}\, e^{-1.75q/kT} \text{ cm}^2/\text{second}, \qquad (13)$$

where q is the electronic charge, k is Boltzman's constant and T is the absolute temperature (kT/q in this expression is in volts). Using (9), (11) and (13), n_s/n_0 was calculated for different values of D, the impurity diffusion coefficient in silicon, at various temperatures. The results are shown in Fig. 4.* On the 1100°C line, points corresponding to various impurities are indicated, using the diffusion coefficients at 1100°C reported by Fuller and Ditzenberger.[5] The broken line shown passing through the antimony point corresponds to the effect of temperature on surface concentration. It corresponds to an activation energy of 4.0 ev for the diffusion coefficient. In general, for a diffusion activation energy greater than the oxidation activation energy of 1.75 ev, lowering the temperature will raise the surface concentration of the redistributed impurity. Only when the activation energy of diffusion of an impurity equals that of the film growth process does the surface concentration become independent of temperature.

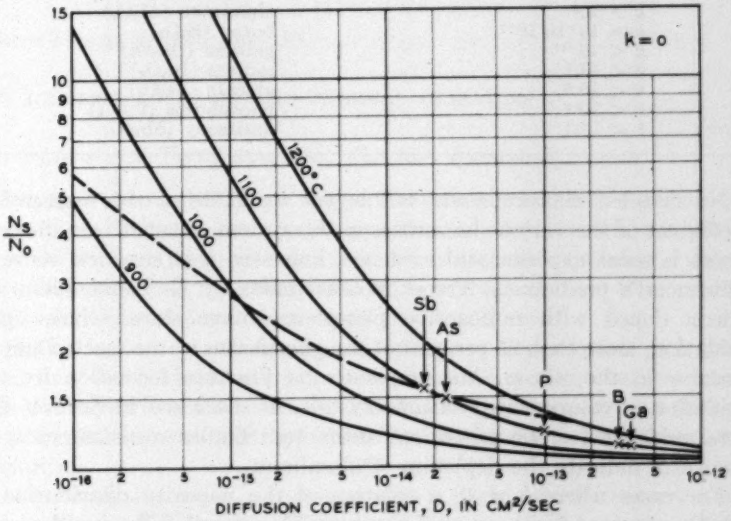


Fig. 4 — Calculated ratios of impurity surface concentration to body concentration (N_s/N_0) produced by parabolic oxidation of silicon with zero segregation coefficient. The arrows indicate the surface concentrations predicted for the impurities indicated at 1100°C. The dashed line corresponds to the surface concentrations predicted for antimony at various temperatures.

* For amorphous silica, $b = 0.44$.

2.3 *Effect of Segregation Coefficient* ($k \neq 0$) *on Impurity Redistribution for Parabolic Film Growth* ($m = 2$)

Based on thermodynamic considerations, Thurmond[6] has calculated the segregation coefficient of various impurites at the interface between silicon and amorphous silicon dioxide. His results are summarized in Table I.

This table indicates that, for most impurities commonly used for doping of silicon, the segregation coefficient is small and should produce pile-up upon oxidation. Furthermore, for these impurities, the assumption of complete rejection by the oxide appears reasonable. For aluminum as an impurity, however, one should obtain depletion near the silicon surface upon oxidation.

TABLE I — SEGREGATION COEFFICIENTS OF VARIOUS IMPURITIES IN THE SYSTEM SILICON-SILICON DIOXIDE

Segregation Coefficient	Impurities
$k > 10^{3}$	aluminum (Al_2O_3)
$k = 10^{3}$ to 10^{-3}	boron (B_2O_3)
$k < 10^{-2}$	gallium (Ga_2O_3)
$k < 10^{-3}$	indium (In_2O_3)
$k < 10^{-3}$	arsenic (As_2O_3)
$k < 10^{-3}$	phosphorus (P_4O_{10})
$k < 10^{-3}$	antimony (Sb_2O_4)

No detailed measurements are as yet available of the segregation coefficient of the various impurities in the system silicon-silicon dioxide. There is some experimental evidence, however, that supports some of Thurmond's predictions. Tracer measurements by E. Tannenbaum on silicon doped with radioactive phosphorus have shown that, upon oxidation, more than 95 per cent of the phosphorus in the reacted silicon remains in the silicon. Measurements on junction formation by the oxidation of compensated silicon crystals, as discussed in Section III, give evidence for the pile-up of antimony. Similar measurements by Doucette indicate the depletion of aluminum.

For cases where $k \neq 0$, a solution of the impurity distribution is readily obtained for the case of parabolic film growth following the same analysis of Section 2.2. The results are as follows:

$$\frac{n_s}{n_0} = \frac{P'}{1 + 2k(P' - 1)} \left\{ 1 + \left[1 - \frac{1 + 2k(P' - 1)}{P'^2} \right]^{\frac{1}{2}} \right\}$$

and

$$\frac{A}{bX} = \frac{1 - k\dfrac{n_s}{n_0}}{\dfrac{n_s}{n_0} - 1}, \tag{14}$$

where

$$P' = 1 + (1 - k)\frac{b^2K}{4D}.$$

It is evident that, by setting $k = 0$, these expressions reduce to those of (9), which is the case of extreme pile-up. For the case of extreme depletion, one sets $k = \infty$ and obtains:

$$\frac{n_s}{n_0} = 0 \tag{15}$$

and

$$\frac{A}{bX} = 2\left[\left(\frac{P+1}{P-1}\right)^{\frac{1}{2}} - 1\right],$$

where P is as defined by (11).

III. JUNCTION FORMATION BY THERMAL OXIDATION

3.1 *System with Two Impurities; Junction Formation*

We have shown that by growing a film on a semiconductor surface one may obtain, under certain conditions, pile-up or depletion of an impurity near the surface. By proper choice of two impurities, one being a donor-type impurity and the other an acceptor-type impurity, it is evident that various junction configurations may be obtained.

In this section we will discuss the conditions under which such junctions may be obtained, and will determine some of their characteristics. Use will be made of the results obtained in the previous sections. Hence, it is implicitly assumed that there are no interactions between the impurities.

In general, the choice of the impurities is based on differences in their segregation coefficient or on differences in diffusion coefficients or on both. In the silicon-silicon dioxide system, for instance, two interesting pairs of impurities are gallium-antimony and phosphorus-aluminum. In the gallium-antimony system, both impurities will pile up with nearly zero segregation coefficient. Due to the smaller diffusion coefficient of antimony, however, it will produce larger pile-up. If this difference exceeds

the initial excess of gallium over antimony in the silicon crystal, a junction is obtained (see Fig. 4). In the phosphorus-aluminum system,[4] on the other hand, phosphorus will pile up while aluminum will be depleted. Here, again, with proper initial compensation of the silicon crystal, a junction may be obtained.

Consider the case of a crystal uniformly doped with a donor-type impurity n_d and an acceptor-type impurity n_a, both with zero segregation coefficient at the semiconductor-film surface. If the crystal is initially p-type (i.e., $n_a > n_d$) the condition for the formation of a *single* junction upon film growth is that $(n_d)_s > (n_a)_s$ and, from (9), this sets the following condition on the initial crystal compensation:

$$\frac{n_d}{n_a} > \frac{P_a + (P_a^2 - 1)^{\frac{1}{2}}}{P_d + (P_d^2 - 1)^{\frac{1}{2}}}. \tag{16}$$

The resulting distribution profiles will be as shown in Fig. 2(a). In a similar fashion, one may set the condition of crystal compensation to obtain two junctions as shown in Fig. 2(b).

To demonstrate some typical characteristics of the obtained junction, we will consider only the simple case where the pile-up of one impurity can be ignored with respect to the pile-up of the other impurity. If n_1 is the impurity that piles up and n_2 is the other, the compensation of the crystal must be such that $n_2 > n_1$, and to obtain a junction one must satisfy the following requirement:

$$\frac{n_2}{n_1} < P_1 + (P_1^2 - 1)^{\frac{1}{2}}. \tag{17}$$

Now an expression will be given for the resulting junction depth y_j, gradient a, and sheet resistance ρ_s by using (6), which describes the distribution profile of the impurity n_1:

$$\frac{y_j}{X} = \frac{bn_1}{n_{1s} - n_1} \ln\left(\frac{n_{1s} - n_1}{n_2 - n_1}\right), \tag{18}$$

$$aX = -\frac{1}{b}(n_2 - n_1)\left(\frac{n_{1s}}{n_1} - 1\right), \tag{19}$$

and

$$\frac{1}{\rho_s X} = \mu q \frac{b}{\frac{n_{1s}}{n_1} - 1}\left[(n_{1s} - n_2) - (n_2 - n_1)\ln\left(\frac{n_{1s} - n_1}{n_2 - n_1}\right)\right], \tag{20}$$

where n_{1s} is the surface concentration of the impurity n_1, and is obtained from (9); μ is the mobility of the carrier of interest; and q is the

electron charge. The right-hand sides of the above equations are independent of time but dependent on temperature alone. For the parabolic film growth, therefore, the junction depth is directly proportional to film thickness, and both junction gradient and sheet resistance are inversely proportional to film thickness. The constants of proportionality, which are the right-hand sides of the above equations, are determined by a single parameter K/D, which is the ratio of the oxidation constant to the diffusion coefficient of the impurity in the semiconductor.

A point of interest that may be mentioned here is that, for a film growth process with well-defined kinetics, it is possible from measurements of resulting junction characteristics (such as junction depth and sheet resistivity) to obtain directly the impurity diffusion constant and its activation energy, and also the carrier mobility in the semiconductor. This is demonstrated in the following section on experimental measurements in the system silicon-silicon dioxide with gallium and antimony as the two impurities.

3.2 *Experimental Results*

All measurements reported here were obtained on silicon with gallium and antimony as the doping impurities. The choice of these two impurities was made for the following reasons. According to Thurmond's calculations (Table I) both impurities should be rejected by the oxide, and hence both should pile up when the silicon surface is oxidized. Furthermore, the diffusion coefficient of antimony is smaller by about two orders of magnitude than that of gallium, and hence should produce stronger pile-up at the surface. Also, the diffusion coefficient of gallium is sufficiently high compared to the oxidation constant that its redistribution can be neglected (see Fig. 4).

We have therefore chosen a silicon crystal that was doped with antimony to a concentration of 4×10^{15} atoms/cm^3 and compensated with gallium to a higher concentration, making the crystal p-type. The crystal was not uniform in resistivity, and small slices with different resistivities, but uniform within a slice, could be obtained from different regions of the crystal. Two sets of slices were used. The first consisted of only one slice with a resistivity of 21 ohm-cm. This was used for the experiment at 1050°C, with the sample being oxidized in oxygen at one atmosphere pressure for the periods of 2, 18 and 90 hours. After each run, the junction depth and sheet resistivity were measured. The surface was then etched off to remove the junction, the sample resistivity was measured and the second oxidation run was carried out. This was repeated until three sets of measurements were made. The second set of slices con-

sisted of two slices of 4 ohm-cm resistivity. These were treated at 1210°C also in oxygen at one atmosphere pressure, one for 1 hour and the other for 16 hours. For each slice, the junction depth and sheet resistance were measured. Furthermore, small mesas were etched out of each slice and the junction breakdown voltage was measured. Using Miller's relations[7] between breakdown voltage and junction gradient, the junction gradient was obtained. Table II gives a summary of the results.

Fig. 5 is a plot of the above data, with the oxide film thickness, junction depth and sheet resistivity for each temperature being plotted versus time of oxidation. The straight lines shown indicate, as predicted, a satisfactory fit of the data to a square root time dependence. In Table II it is also seen that the junction gradient has approximately the same time dependence. According to (18), (19) and (20), the parameters y_j/X, aX and $\rho_s X$ should be functions of the single ratio K/D, which is only dependent on temperature.

In Table III, the experimental values are given for the parameters y_j/X, aX, $\rho_s X$ and $X^2/t = K$ at the two temperatures of 1210 and 1050°C. Using (18), (19) and (20), it is now possible to substitute the experimental values of the above parameters to obtain the ratio K/D and the diffusion coefficient D for the two temperatures and compare with published data. In principle, the same value of the ratio K/D should be obtained at any one temperature from all three measurements of y_j/X, aX and $\rho_s X$. Such calculations were carried out and the results are also shown in Table III. At 1210°C, K/D was found to be 0.18 and 0.14, from the junction depth and the junction gradient measurements respectively. To obtain K/D from (20) and the measured sheet resistance parameter, $\rho_s X$, one finds a strong dependence of K/D on the value assigned to the mobility μ_n. We have instead substituted the values of K/D of 0.18 and 0.14, obtained from junction depth and junction gradient measurements, in (20) and obtained values of μ_n corresponding to the measured parameter $\rho_s X$. The values of μ_n ob-

TABLE II — JUNCTION CHARACTERISTICS OF OXIDIZED P-TYPE COMPENSATED SILICON WITH GALLIUM-ANTIMONY AS IMPURITIES*

T, °C	t, hours	ρ, ohm-cm	X, A°	Y_j, A°	ρ_s, ohm/sq	V_B, volts	a, cm^{-4}
1050	2	21	1250	2800	2480	—	—
1050	18	21	3800	10300	880	—	—
1050	90	21	10500	26400	414	—	—
1210	1	3.65	2100	15200	2000	55	5.5×10^{20}
1210	16	3.65	8400	73700	520	85	1.55×10^{20}

* Antimony concentration $\sim 4 \times 10^{18}$ atoms/cc.

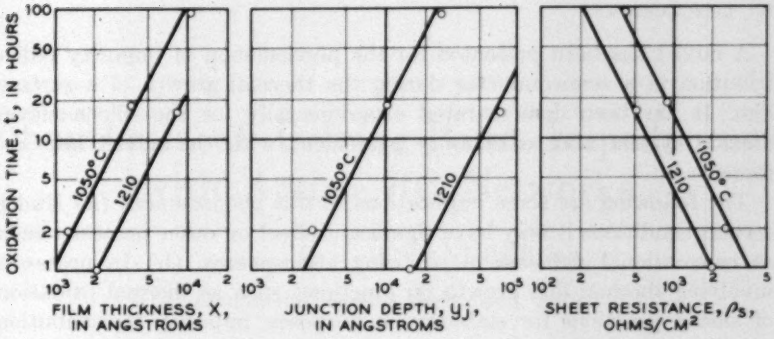


Fig. 5 — Experimental data showing the variations with time of the oxide film thickness, the junction depth and the sheet resistance, for gallium-antimony doped silicon. The straight lines correspond to $t^{\pm\frac{1}{2}}$ dependence.

tained were 120 and 115 cm²/volt-seconds respectively. Similarly, the data were analyzed and the results are given in Table III. It is to be noted that the obtained values of electron mobility μ_n of 97–120 cm²/volt-second at 4×10^{18}/cm³ concentration are in satisfactory agreement with published data (see, for instance, Conwell's review article[8]). From the calculated values of K/D and the measured $K = X^2/t$, the diffusion coefficients of antimony in silicon at 1050 and 1210°C were obtained and are given in Table III. Corresponding values from the literature[5] are also given for comparison.

The self-consistency of the above results and their fair agreement with published data are quite satisfactory considering the many simplifying assumptions on which the analysis was based.

TABLE III — CONSTANTS CALCULATED FROM MEASURED PARAMETERS

Measured					Calculated		
T, °C	Y_j/X	aX, cm^{-1}	$\rho_s X$, ohm-cm	K, cm²/sec	K/D	μ_n, cm²/volt-sec	D, cm²/sec
1210	8.0	1.2×10^{16}	0.042	1.2×10^{-12}	0.18 0.14	115 120	(a) 7×10^{-12}*
1050	2.6	—	0.037	2.6×10^{-14}	5.8	97	(b) 4.5×10^{-15}*

* Published values (Fuller and Ditzenberger[5]): (a) 2.5×10^{-12}; (b) 5×10^{-15}.
μ_p = 40 cm²/volt-sec.
$C_{sb} = 4 \times 10^{18}$ atoms/cc.

IV. CONCLUSIONS

A model has been presented for the phenomenon of impurity redistribution in a semiconductor during the thermal growth of a surface film. It has been demonstrated experimentally for the silicon-silicon dioxide system and satisfactory agreement with the model was obtained.

The following are some implications of this phenomenon: (a) Under certain conditions it may have significant effect on other processes such as conventional diffusion in oxidizing atmospheres. (b) In processes involving thermal film growth on junctions, such as thermal oxidation of silicon junctions for stabilization purposes, impurity redistribution effects can be significant under certain conditions and must be taken into consideration. (c) The redistribution phenomenon provides an experimental tool for determinations of some physical constants such as segregation coefficients, diffusion constants and mobilities. (d) One can make junctions by this method. However, heavily compensated starting materials are needed, which often poses a control problem.

REFERENCES

1. Ligenza, J. R. and Spitzer, W., to be published.
2. Atalla, M. M., in *Properties of Elemental and Compound Semiconductors*, Interscience, New York, 1960, p. 163.
3. Ligenza, J. R., to be published.
4. Doucette, E. I., Cooper, N. W. and Mehnert, R. A., to be published.
5. Fuller, C. S. and Ditzenberger, J. A., J. Appl. Phys., **27**, 1956, p. 544.
6. Thurmond, C. D., in *Properties of Elemental and Compound Semiconductors*, Interscience, New York, 1960, p. 121.
7. Miller, S. L., private communication.
8. Conwell, E., Proc. I.R.E., **46**, 1958, p. 1281.

Synthesis of Driving-Point Impedances with Active *RC* Networks

By I. W. SANDBERG

(Manuscript received December 4, 1959)

A general method is presented for synthesizing driving-point impedances using RC networks and active elements. The procedure realizes any real rational driving-point function and leads to rather simple structures. Only one active device, a negative-impedance converter, is required. The synthesis of biquadratic impedance functions is considered in detail.

I. INTRODUCTION

It is often desirable to avoid the use of magnetic elements in synthesis procedures, since resistors and capacitors are more nearly ideal elements and are usually cheaper, lighter and smaller. This is especially true in control systems in which, typically, exacting performance is required at very low frequencies. The rapid development of the transistor has provided the network synthesist with an efficient low-cost active element and has stimulated considerable interest in active *RC* network theory during the past decade.[1,2,3,4,5]

The present paper considers the active *RC* synthesis of driving-point impedances. Transfer functions are not treated directly, but are covered at least in principle, since it is always possible, and indeed sometimes convenient, to reduce the synthesis of transfer functions to the synthesis of two-terminal impedances.

It is now well known that any driving-point impedance function expressible as a real rational fraction in the complex frequency variable can by synthesized as an active *RC* network requiring only one ideal active element. Two proofs of this result are already in the literature.[6,7]* The present paper provides a third proof, although its main objective is to present a new and more practical realization network. The synthesis

* Another proposed proof[8] is in fact concerned only with those impedance functions which are positive on some section of the negative-real axis of the complex frequency plane.

technique is similar to that used in the author's first proof,[6] and considerably simpler than those used in the later paper by Kinariwala.[7]

Like the technique previously presented by the author,[6]* the present technique yields the values of the required network elements explicitly and directly, without the solution of simultaneous equations. The general procedure yields networks that contain more capacitors than the absolute minimum required; in return for this, however, it yields structures which, unlike those of Kinariwala, do not need balanced amplifiers or complex resistance networks.†

An alternative procedure applicable to a wide class of biquadratic impedance functions is also described. This procedure leads to structures requiring only two capacitors.

The synthesis techniques presented in this paper are based on networks employing a type of impedance converter that is a generalization of the negative-impedance converter. The converter concept is introduced in a general way in order to properly orient the reader.

II. THE IDEAL IMPEDANCE CONVERTER

Consider a two-port network terminated by an impedance $Z_T(s)$ as shown in Fig. 1. The input impedance at port 1 is

$$Z_1(s) = h_{11} - \frac{h_{12}h_{21}Z_T(s)}{1 + h_{22}Z_T(s)}, \tag{1}$$

where the hybrid parameters are functions of the complex frequency variable defined by

$$\begin{aligned} E_1 &= h_{11}I_1 + h_{12}E_2\,, \\ I_2 &= h_{21}I_1 + h_{22}E_2\,. \end{aligned} \tag{2}$$

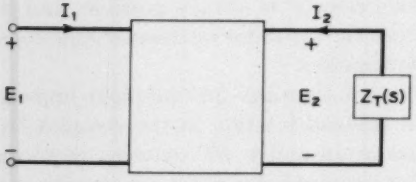


Fig. 1 — Two-port network.

* The particular configuration involved is essentially the one mentioned in Section 3.2 of this paper.

† A more practical procedure in Ref. 7 treats a restricted class of driving-point functions and employs a passive *RC* two-port network terminated by a negative *RC* impedance. Other restricted realization techniques are presented in Ref. 8.

For the ideal impedance converter we require that for every $Z_T(s)$

$$Z_1(s) = K(s)Z_T(s), \tag{3}$$

where $K(s)$ is a predetermined fixed function of the complex frequency variable. In order that (1) and (3) be compatible,

$$\begin{aligned} h_{11} &= h_{22} = 0, \\ K(s) &= -h_{12}h_{21}. \end{aligned} \tag{4}$$

An ideal impedance converter, therefore, is a two-port network with a hybrid parameter matrix of the form

$$\begin{bmatrix} 0 & h_{12} \\ h_{21} & 0 \end{bmatrix}. \tag{5}$$

It follows that the impedance at port 2 with the termination $Z_T(s)$ connected to port 1 is

$$Z_2(s) = \frac{Z_T(s)}{K(s)}. \tag{6}$$

A "controlled-source" representation of the unbalanced ideal impedance converter is given in Fig. 2.

The negative-impedance conveter ($h_{12}h_{21} = 1$) has been heavily exploited as a synthesis tool in active RC network theory. The synthesis techniques presented in this paper are based on networks employing a more general type of converter characterized by

$$\begin{bmatrix} 0 & 1 \\ \dfrac{Z_3}{Z_4} & 0 \end{bmatrix}, \tag{7}$$

where Z_3/Z_4 is the ratio of two RC driving-point impedances. This hybrid parameter matrix can be realized with either of the idealized cir-

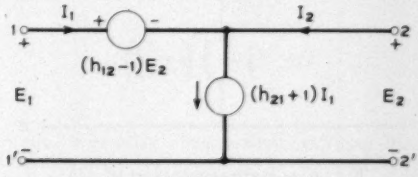


Fig. 2—Representation of the ideal impedance converter.

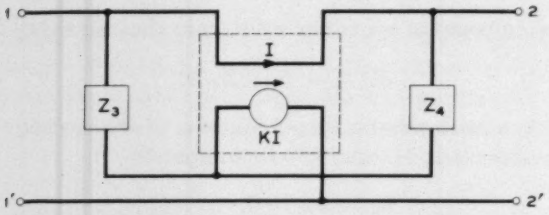


Fig. 3 — Idealized realization of the required converter.

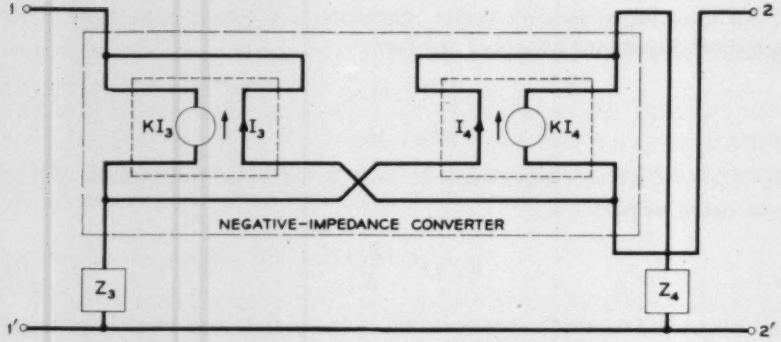


Fig. 4 — Alternate idealized realization of the converter.

cuits shown in Figs. 3 and 4.* The smaller rectangles enclose "infinite gain" current amplifiers. The realization of Fig. 3 is a modification of Larky's idealized current-inversion negative-impedance converter.[9] The larger rectangle in Fig. 4 encloses Linvill's well-known idealized voltage-inversion negative-impedance converter.[10] Hence, (7) can be realized with Linvill's negative-impedance converter and two *RC* driving-point impedances.

A controlled-source representation of (7) is given in Fig. 5. Transistor realizations are discussed in the Appendix.

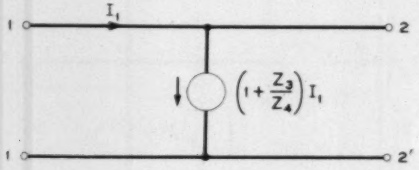


Fig. 5 — Representation of (7).

* Many other realizations are possible.

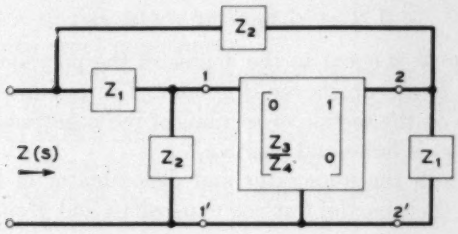


Fig. 6 — Realization one-port.

III. SYNTHESIS OF DRIVING-POINT IMPEDANCES

The driving-point impedance of the one-port network shown in Fig. 6 is*

$$Z(s) = \frac{Z_4 - Z_3}{\frac{Z_4}{Z_1} - \frac{Z_3}{Z_2}}. \tag{8}$$

The impedance function that is to be synthesized is a real rational fraction in the complex frequency variable

$$Z(s) = \frac{P(s)}{Q(s)}. \tag{9}$$

The synthesis consists of identifying each of the four parameters Z_1, Z_2, Z_3 and Z_4 with a two-terminal RC impedance function. The presence of negative signs in (8) suggests an approach similar to techniques previously proposed for the synthesis of transfer functions.[3,5]

Assume that the prescribed function $Z(s)$ is positive on at least one section of the negative-real axis of the complex frequency plane. Suppose we write

$$Z(s) = \frac{\frac{P(s)}{\prod_{i=1}^{N}(s + \sigma_i)}}{\frac{Q(s)}{\prod_{i=1}^{N}(s + \sigma_i)}}, \tag{10}$$

where

* This expression can be readily obtained by applying Blackman's equation[11] to the network that results when the converter is replaced by its controlled source representation. The required return differences are computed with respect to the quantity $1 + (Z_3/Z_4)$. A detailed study of Blackman's equation led to the discovery of the realization networks presented in this paper.

$$0 \leqq \sigma_1 < \sigma_2 < \sigma_3 \cdots < \sigma_N ,$$

and the number N is equal to the degree of the polynomial $P(s)$ or of the polynomial $Q(s)$, whichever is greater. The points $-\sigma_i$ are chosen to lie anywhere on the section or sections of the negative-real axis where the function $Z(s)$ is finite and positive.*

We replace both the numerator and denominator of the right-hand side of (10) by their partial fraction expansions and group the resulting terms to obtain

$$Z(s) = \frac{\dfrac{P_1}{Q_1} - \dfrac{P_2}{Q_2}}{\dfrac{P_3}{Q_1} - \dfrac{P_4}{Q_2}} \tag{11}$$

where

$$Q_1 Q_2 = \prod_{i=1}^{N} (s + \sigma_i)$$

and P_1/Q_1 , P_2/Q_2 , P_3/Q_1 , and P_4/Q_2 are each RC driving-point impedances. The function $Z(s)$ is expressible in this form since:

(a) Each of the two partial fraction expansions possesses real residues of like sign at any particular pole.

(b) All RC driving-point impedances are expressible in the form

$$R_\infty + \sum_{k=1}^{M} \frac{a_k}{s + b_k},$$

where R_∞ , a_k and b_k are nonnegative.

Equation (11) is equivalent to

$$Z(s) = \frac{\dfrac{P_1}{Q_1} - \dfrac{P_2}{Q_2}}{\dfrac{P_3 + RQ_1}{Q_1} - \dfrac{P_4 + RQ_2}{Q_2}}. \tag{12}$$

We identify the impedance parameters of (8) as follows:

$$\begin{aligned} Z_4 &= \frac{P_1}{Q_1}, & Z_3 &= \frac{P_2}{Q_2}, \\ Z_1 &= \frac{P_1}{P_3 + RQ_1}, & Z_2 &= \frac{P_2}{P_4 + RQ_2}. \end{aligned} \tag{13}$$

* The choice of the σ_i's influences the spread of element values and the sensitivity of the driving-point impedance to variations in the active and passive elements.

According to the qualifications appended to (11), the functions Z_3 and Z_4 are RC driving-point impedances.

Consider the function Z_1. The degree of the polynomial Q_1 is at least as great as the degree of P_3. Note that, as the parameter R assumes nonnegative real values that increase from zero to infinity, the poles of Z_1 move from the zeros of P_3 to the zeros of Q_1. Recall that the ratio of polynomials P_1/Q_1 is an RC driving-point impedance. Consequently, it is always possible to find a finite value of the parameter R for which the poles and zeros of Z_1 interlace properly.

A similar argument shows that Z_2 also can always be made an RC driving-point impedance. Note that $Z(s)$ need not be a positive-real function.

3.1 *An Example*

Let

$$Z(s) = \frac{s^2 + s + 1}{s^2 + s + 2}. \tag{14}$$

This function is positive on the entire negative-real axis. We choose $\sigma_1 = 1$, and $\sigma_2 = 2$. From (12),

$$Z(s) = \frac{\dfrac{s+2}{s+1} - \dfrac{3}{s+2}}{\dfrac{s(1+R)+3+R}{s+1} - \dfrac{Rs+4+2R}{s+2}}. \tag{15}$$

Employing (13),

$$\begin{aligned} Z_4 &= 1 + \frac{1}{s+1}, & Z_3 &= \frac{3}{s+2}, \\ Z_1 &= \frac{s+2}{s(1+R)+3+R}, & Z_2 &= \frac{3}{Rs+4+2R}. \end{aligned} \tag{16}$$

The set of impedances is realizable for $R \geqq 1$. Choose $R = 1$ so that

$$Z_1 = \frac{1}{2}, \qquad Z_{12} = \frac{3}{s+6}. \tag{17}$$

The corresponding network is shown in Fig. 7.

Note that Z_3 and Z_4 can be scaled by the same constant without affecting the impedance $Z(s)$. In a practical design, this degree of freedom would be utilized in order to optimize some figure of merit such as the sensitivity function. The choice of the σ_i's would be similarly influenced.

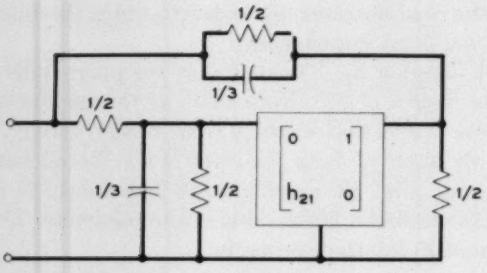


Fig. 7 — Realization of (14).

3.2 *An Alternate Realization*

The class of functions treated in this section can also be synthesized with the network given in Fig. 8. The input impedance of this structure is

$$Z(s) = \frac{Z_8 Z_6 - Z_7 Z_5}{Z_8 - Z_7}. \tag{18}$$

By employing arguments similar to those already discussed, it can be shown that the value of a nonnegative real parameter R can be chosen to ensure that the following set of impedances is realizable:*

$$\begin{aligned} Z_8 &= \frac{P_3 + RQ_1}{Q_1}, \qquad Z_7 = \frac{P_4 + RQ_2}{Q_2}, \\ Z_6 &= \frac{P_1}{P_3 + RQ_1}, \qquad Z_5 = \frac{P_2}{P_4 + RQ_2}. \end{aligned} \tag{19}$$

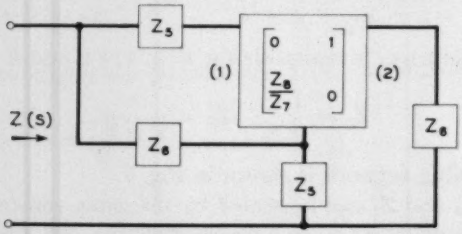


Fig. 8 — Alternate realization one-port.

* A realizable set of impedances can also be obtained in other ways.

This circuit has the disadvantage of requiring a "floating" two-port active network.

3.3 *Restrictions on* $Z(s)$

The methods of synthesis presented in the previous sections are not applicable to functions that are nonpositive on the entire negative-real axis. This is a significant theoretical restriction, since positive-real functions, for example, need not possess the required property. In particular, all reactance functions must be excluded.

The difficulty mentioned above can be circumvented in several ways by modifications of the synthesis technique. Suppose that the prescribed impedance $Z(s)$ is nonpositive on the entire negative-real axis. The function

$$Z'(s) = Z(s) - \frac{a_0}{s + b_0} \qquad a_0 > 0, \quad b_0 \geqq 0 \tag{20}$$

must, however, be positive on one section of the negative-real axis. It can therefore be synthesized by the previously discussed procedure. The impedance $Z(s)$ is obtained by connecting an RC impedance $a_0/(s + b_0)$ in series with the resulting network. An alternative procedure on an admittance basis also applies, the network being modified at the input terminals by the parallel connection of an RC impedance

$$c_0 + \frac{d_0}{s} \qquad d_0 > 0, \quad c_0 \geqq 0.$$

Both methods usually necessitate a larger number of passive components than would be required for the synthesis of $-Z(s)$. For this reason it may be more desirable to employ a negative-impedance converter terminated by $-Z(s)$.

3.4 *Sufficiency of One Active Element or One Negative-Impedance Converter*

Since the realization of the converter requires only one active element or only one Linvill-type negative-impedance converter, the preceding discussion constitutes a proof of

Theorem: Any driving-point impedance function, expressible as a real rational fraction in the complex frequency variable, can be synthesized with a network containing only resistors, capacitors and either a single ideal active element or a single ideal negative-impedance converter.

Note that it is theoretically possible to synthesize impedance functions which approach infinity as any integral power of the frequency

variable. Obviously, these functions can be realized by actual networks only over the frequency band where the active two-port is essentially characterized by the controlled-source model of Fig. 5.*

IV. BIQUADRATIC SYNTHESIS

The synthesis of biquadratic impedance functions merits special attention. The network associated with a function of this type, previously considered as an example, requires a total of four capacitors (two in the converter). Two of the capacitors can be eliminated by employing an alternative technique based on the network of Fig. 6 with Z_1 and Z_2 replaced by resistors R_1 and R_2. The structure becomes an impedance converter imbedded in a simple resistance network.

From (8),

$$\frac{Z_3}{Z_4} = \frac{R_2(Z - R_1)}{R_1(Z - R_2)}. \tag{21}$$

The biquadratic function $Z(s)$ is given by

$$Z(s) = \frac{(s - z_1)(s - z_2)}{(s - p_1)(s - p_2)}. \tag{22}$$

It is required that R_1 and R_2 be chosen so that Z_3 and Z_4 are RC impedances. Assume that each of the impedances Z_3 and Z_4 is to be realizable as a resistor in series with a parallel combination of a resistor and capacitor as shown in Fig. 9. This structure is sufficient to realize the most general first-degree RC impedance function. The ratio of these two functions is given by

$$\frac{Z_3}{Z_4} = K\frac{(s + \alpha_1)(s + \alpha_2)}{(s + \alpha_3)(s + \alpha_4)}, \tag{23}$$

where, from (21) and (22),

$$K = \frac{R_2(1 - R_1)}{R_1(1 - R_2)}. \tag{24}$$

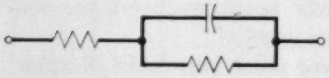


Fig. 9 — Structural form of Z_3 or Z_4.

* It must be remembered that this model will ordinarily be inadequate for stability analyses. For this purpose it must be modified to be valid in the frequency range where the significant active and passive parasitic parameters are influential.

Four possibilities exist for the pole-zero pattern of Z_3/Z_4 as shown in Fig. 10. The zeros in Fig. 10(a) and the poles in Fig. 10(b) may occur with multiplicity two.

Assume tentatively that $Z(s)$ has complex conjugate poles and zeros. The function $Z(\sigma)[-\infty \leqq \sigma \leqq \infty]$, where $s = \sigma + j\omega$, is nonnegative and approaches unity at both extremes of the argument. Since the function is the ratio of two second-degree polynomials in σ, only two points

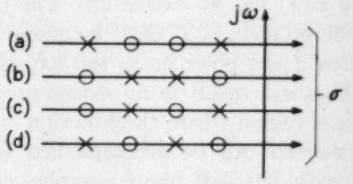


Fig. 10 — Permissible pole-zero patterns for Z_3/Z_4.

of intersection with a horizontal line are possible. When such an intersection occurs, the intersecting points will be separated by an extremum of the function. Hence, if $Z(s)$ exhibits at least one extremum on the negative-real axis, the parameters R_1 and R_2 can be chosen to provide a pole-zero pattern for Z_3/Z_4 of the type shown in Figs. 10(a) or (b). Since, for both type (a) and type (b), $(1 - R_1)$ and $(1 - R_2)$ have the same sign, the impedances Z_3 and Z_4 would be realizable.

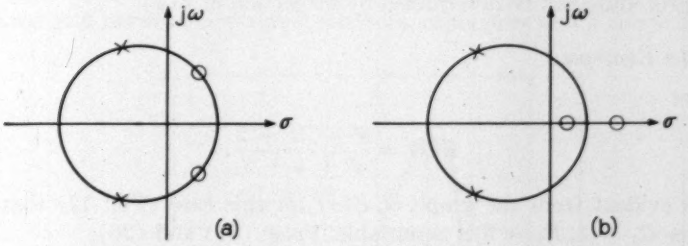


Fig. 11 — Construction of the extremum points of $Z(\sigma)$.

Consider the pole-zero diagram for $Z(s)$. It can easily be shown that the circle passing through the poles and zeros [Fig. 11(a)] will intersect the real axis at the two points where $Z(\sigma)$ has an extremum. Consequently, any open-circuit or short-circuit stable biquadratic impedance function with complex conjugate poles and zeros can be synthesized. Two permissible limiting cases exist when the zeros or poles or both the zeros and poles occur with multiplicity two on the real axis. Another

permissible case occurs when the zeros are at infinity and the complex conjugate poles are in the left-half plane.

Suppose that $Z(s)$ has distinct real zeros and complex conjugate poles. It can easily be shown that the circle centered on the real axis at

$$\sigma_0 = \frac{\sigma_p^2 + \omega_p^2 - z_1 z_2}{2\sigma_p - (z_1 + z_2)} \qquad p_{1,2} = \sigma_p \pm j\omega_p \tag{25}$$

and passing through the poles [Fig. 11(b)] will intersect the real axis at the two points where $Z(\sigma)$ has an extremum. The circle must pass between the zeros. It follows from (25) that the center of the circle will lie to the left of the zeros if the poles lie to the left of the point midway between the zeros. This will result in an extremum of $Z(\sigma)$ located to the left of the poles in a region where the function is positive. It follows directly that the synthesis can be accomplished for any biquadratic impedance function with left-half plane complex conjugate poles and real zeros, where the poles are located to the left of the point midway between the zeros.*

The steps in the biquadratic synthesis procedure are: (a) choose R_1 and R_2 so that Z_3 and Z_4 are realizable, and (b) from (21), identify Z_3 and Z_4. The permissible values of R_1 and R_2 can be determined by inspection of $Z(\sigma)$.

It should be noted that this procedure is not limited to the two broad classes of functions considered above. The synthesis can obviously be accomplished if step (a) can be carried out. Hence the applicability of the procedure can be determined by inspection of $Z(\sigma)$.

4.1 *An Example*

Let

$$Z(s) = \frac{s^2 - 3s + 2}{s^2 + s + 1}. \tag{26}$$

It is evident from the graph of $Z(\sigma)$ for this case (Fig. 12) that the choice $R_1 = 4$, $R_2 = 6$ is acceptable. From (21) and (26),

$$\begin{aligned} \frac{Z_3}{Z_4} &= \frac{6}{4}\frac{Z - 4}{Z - 6} = \frac{9}{10}\frac{s^2 + \frac{7}{3}s + \frac{2}{3}}{s^2 + \frac{3}{5}s + \frac{1}{5}} \\ &= \frac{9}{10}\frac{(s + \frac{1}{3})(s + 2)}{(s + \frac{1}{5})(s + 1)}. \end{aligned} \tag{27}$$

Two possibilities exist for the pair of impedances Z_3 and Z_4:

* Biquadratic admittance functions with distinct negative-real poles can be realized with a negative-impedance converter as the difference of two RC admittances. For this reason, such functions are not considered in detail here.

$$Z_3 = \frac{9}{10}\frac{(s+2)}{(s+1)}, \qquad Z_4 = \frac{(s+\frac{4}{3})}{(s+\frac{1}{3})}, \tag{28}$$

or

$$Z_3 = \frac{9}{10}\frac{(s+2)}{(s+\frac{4}{3})}, \qquad Z_4 = \frac{(s+1)}{(s+\frac{1}{3})}. \tag{29}$$

4.2 *Synthesis on the Basis of the Network Shown in Fig. 8*

The synthesis technique presented in this section can be extended to apply to the network of Fig. 8 by comparing (8) and (18) and identifying

$$\begin{aligned} Z_7 &= \frac{R_1}{R_2} Z_3, \qquad & Z_8 &= Z_4, \\ Z_6 &= R_1, & Z_5 &= R_2. \end{aligned} \tag{30}$$

V. CONCLUSION

A general method of synthesizing driving-point impedances has been presented. An impedance converter is required that can be realized by modifying Larky's current-inversion negative-impedance converter. An alternate realization employs Linvill's voltage-inversion negative-impedance converter and two *RC* impedances. This realization leads to the result that any driving-point impedance function, expressible as a real rational fraction in the complex frequency variable, can be synthesized as a network containing resistors, capacitors and a single nega-

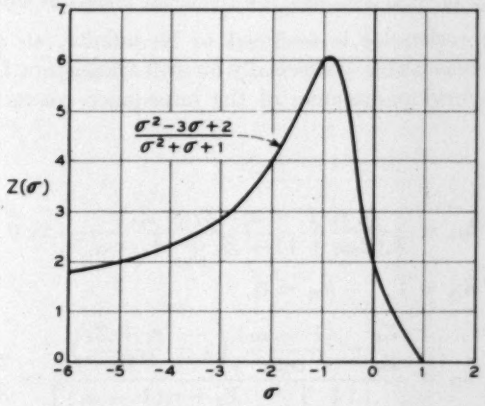


Fig. 12 — Graph of $Z(\sigma)$ for a biquadratic function.

tive-impedance converter. The technique does not require the synthesis of two-port *RC* or *n*-port balanced resistor networks, and leads to the direct determination of the required two-terminal *RC* elements.

The synthesis of biquadratic impedance functions has been given special attention, resulting in structures employing the minimum number of capacitors and a moderate number of resistors. The procedure is applicable to a wide class of functions including, in particular, all open-circuit or short-circuit stable impedances with complex conjugate poles and zeros.

VI. ACKNOWLEDGMENT

The author is grateful to B. McMillan and to P. M. Dollard for their constructive criticism and advice.

APPENDIX

Transistor Circuit Analysis

An approximate model for the transistor in the range where its parameters are essentially independent of frequency is shown in Fig. 13

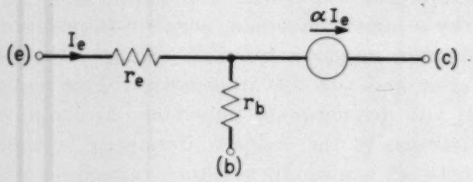


Fig. 13 — Approximate low-frequency transistor model.

The collector resistance is assumed to be infinite, an approximation which is often reasonable — especially for drift transistors. In terms of this model, the hybrid parameters of the impedance converter circuits of Fig. 14 are

Fig. 14(a):

$$\begin{aligned} h_{11} &= \frac{r_1(1-\alpha_1)(Z_3+Z_4)}{Z_4(\beta_2\alpha_1+1)+Z_3+(1-\alpha_1)r_1} \simeq 0, \\ h_{12} &= 1, \qquad h_{22} = 0, \\ h_{21} &= \frac{\dfrac{Z_3}{Z_4} - \dfrac{(1-\alpha_1)}{\alpha_1\beta_2}\left(1+\dfrac{r_1+Z_3}{Z_4}\right)}{1+\dfrac{1}{\alpha_1\beta_2}\left[1+\dfrac{Z_3+r_1(1-\alpha_1)}{Z_4}\right]} \simeq \frac{Z_3}{Z_4}, \end{aligned} \tag{31}$$

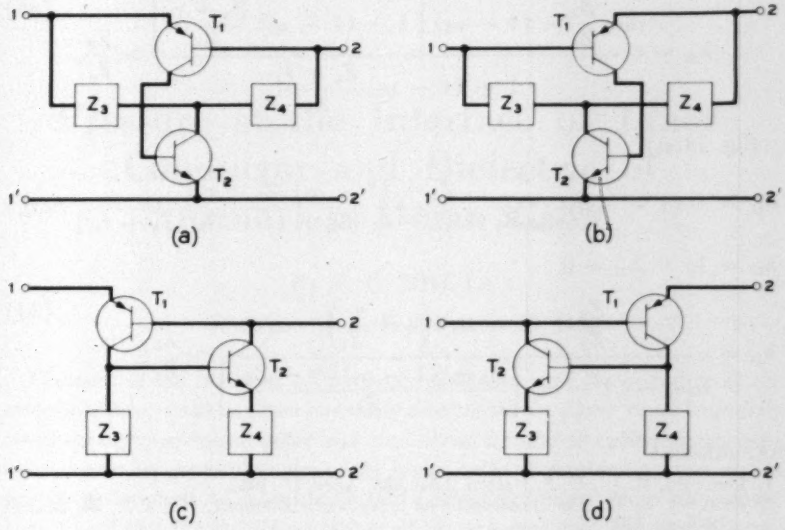


Fig. 14 — Transistor realizations of the converter.

where

$$r = r_b + \frac{r_e}{1-\alpha} \qquad \beta = \frac{\alpha}{1-\alpha}.$$

Fig. 14(b):

$$\begin{aligned} h_{11} &= \frac{-r_1(1-\alpha_1)(Z_3+Z_4)}{\alpha_1\beta_2 Z_4 - (1-\alpha_1)(r_1+Z_3+Z_4)} \simeq 0, \\ h_{12} &= 1, \qquad h_{22} = 0, \\ h_{21} &= \frac{\dfrac{Z_3}{Z_4} + \dfrac{1}{\alpha_1\beta_2}\left[1 + \dfrac{Z_3 + r_1(1-\alpha_1)}{Z_4}\right]}{1 - \dfrac{(1-\alpha_1)}{\alpha_1\beta_2}\left[1 + \dfrac{r_1+Z_3}{Z_4}\right]} \simeq \frac{Z_3}{Z_4}. \end{aligned} \tag{32}$$

Fig. 14(c):

$$h_{11} = (1-\alpha_1)r_1 \simeq 0,$$

$$h_{12} = 1, \qquad h_{22} = 0,$$

$$h_{21} = \frac{\alpha_1\alpha_2 \dfrac{Z_3}{Z_4} - (1-\alpha_1)\left[1 + (1-\alpha_2)\dfrac{Z_3 + r_2}{Z_4}\right]}{1 + (1-\alpha_2)\dfrac{Z_3 + r_2}{Z_4}} \simeq \frac{Z_3}{Z_4}. \tag{33}$$

Fig. 14(d):

$$h_{11} = -(1-\alpha_1)r_1 \frac{Z_3 + (1-\alpha_2)(Z_4 + r_2)}{\alpha_1\alpha_2 Z_4 - (1-\alpha_1)[Z_3 + (1-\alpha_2)(Z_4 + r_2)]} \simeq 0,$$

$$h_{12} = 1, \qquad h_{22} = 0,$$

$$h_{21} = \frac{\dfrac{Z_3}{Z_4} + (1-\alpha_2)\left(1 + \dfrac{r_2}{Z_4}\right)}{\alpha_1\alpha_2 - (1-\alpha_1)\left[\dfrac{Z_3 + (1-\alpha_2)(r_2 + Z_4)}{Z_4}\right]} \simeq \frac{Z_3}{Z_4}. \tag{34}$$

REFERENCES

1. Dietzold, R. L., U. S. Patent 2,549,065, April 17, 1951.
2. Bangert, J. T., The Transistor as a Network Element, B.S.T.J., **33,** March 1945, p. 329.
3. Linvill, J. G., RC Active Filters, Proc. I.R.E., **42,** March 1954, p. 555.
4. Horowitz, I. M., Synthesis of Active RC Transfer Functions, Research Report R-507-56, PIB-437, Microwave Research Institute, Polytechnic Inst. of Brooklyn, November 1956.
5. Yanagisawa, T., RC Active Networks Using Current-Inversion-Type Negative-Impedance Converters, Trans. I.R.E., **CT-4,** September 1957, p. 140.
6. Sandberg, I. W., Active RC Networks, Research Report R-662-58, PIB-590, Microwave Research Institute, Polytechnic Inst. of Brooklyn, May 1958.
7. Kinariwala, B. K., Synthesis of Active RC Networks, B.S.T.J., **38,** September 1959, p. 1269.
8. DeClaris, N., Synthesis of Active Networks — Driving-Point Functions, I.R.E. Nat. Conv. Rec., 1959, p. 23.
9. Larky, A. I., Negative-Impedance Converters, Trans. I.R.E., **CT-4,** September 1957, p. 124.
10. Linvill, J. G., Transistor Negative-Impedance Converters, Proc. I.R.E., **41,** June 1953, p. 725.
11. Bode, H. W., *Network Analysis and Feedback Amplifier Design,* D. Van Nostrand Co., New York, 1956, p. 67.
12. Horowitz, I. M., Optimization of Negative-Impedance Conversion Methods of Active RC Synthesis, Trans. I.R.E., **CT-6,** September 1959.

Pressure at the Interface of Inner Conductor and Dielectric of Armorless Ocean Cable

By P. G. BHUTA

(Manuscript received February 8, 1960)

Pressure at the interface of the inner conductor and the dielectric of the armorless ocean cable is experimentally determined for plane stress boundary conditions. Polyethylene outer and steel inner concentric cylinders are considered as a mechanical model of the ocean cable. Equations are derived to give the pressure at the interface for plane stress and plane strain conditions. It is found that the theoretical values of the pressure calculated from equations of elasticity are in good agreement with experimental values, even when one of the materials is polyethylene, which has a nonlinear stress-strain relationship in simple tension or compression. From previously measured values of bulk and Young's moduli, it is shown that, for an elastic analysis, Poisson's ratio for polyethylene is very nearly 0.5. An experimental verification of the plane stress solution given in this paper confirms the foregoing value.

I. INTRODUCTION

In the process of designing new armorless coaxial ocean cables it is important to know the pressure at the interface of the inner conductor and the polyethylene dielectric. Determination of pressure at the interface would enable one to study the decay of tension, in a cable-laying machine, from the strength member to the holding surface, and to investigate the pressure effects on the resistance and capacitance of the cable.

A suitable mechanical model of the armorless ocean cable is a polyethylene outer cylinder concentric with a steel inner cylinder. The cable is actually subjected, in three dimensions, to a pressure loading that varies with the depth of the ocean bottom. However, for a section of the cable on a relatively flat ocean bottom one may assume that the cable is subjected to a uniform external pressure.

One may obtain an expression for the pressure at the interface from a two-dimensional model (i.e., for plane stress or plane strain) using the theory of elasticity. However, the result is questionable, since polyethylene has a nonlinear stress-strain relationship in simple tension or compression. Moreover the value of Poisson's ratio for polyethylene, μ_p, ranges from 0.3 to 0.5, and it is not possible to determine μ_p accurately by a simple experiment. The value of the pressure at the interface, as obtained from such an equation, depends considerably upon the particular value of μ_p that is used. Hence, an experimental determination of the pressure at the interface is necessary, to investigate whether an elastic analysis is valid when one of the materials is polyethylene and to establish the value of Poisson's ratio for polyethylene.

The pressure at the interface of the model is determined experimentally for plane stress boundary conditions. Equations are derived, based on the two-dimensional theory of elasticity, to give the pressure at the interface of the model for plane stress and plane strain conditions. It is found that the theoretical values of the pressure calculated from equations of elasticity are in good agreement with experimental values. From previously measured values of bulk and Young's moduli, it is shown that, for an elastic analysis, Poisson's ratio for polyethylene is very nearly 0.5. For this reason and since the area of polyethylene is much greater than that of steel, an extension of the two-dimensional plane strain solution to a three-dimensional case of uniform pressure by the method of superposition would give the pressure at the inner conductor and dielectric interface to be equal, or at most slightly greater, than the external pressure.

II. EXPERIMENTAL MODEL AND PROCEDURE

Initially, the experimental model consisted of a cold-drawn steel tubing 30 inches long, 1.254 inches in outer diameter and 0.083 inch in wall thickness. Two SR-4 strain gages of type A7 were cemented on a 0.010-inch-thick steel ring, which fitted snugly inside the steel tubing shown in Fig. 1. The steel ring containing the gages to which long leader wires were soldered was cemented in the tubing so that the it was at the same distance from either end. The gages were far enough from the ends to eliminate end effects.

The tubing was subjected to an external hydrostatic pressure in increments of 500 psi up to 4500 psi when a maximum strain of approximately 1000 microinches per inch, which is well within the elastic limit for the tubing, was recorded. The pressure was reduced to zero, and upon increasing the pressure again, the same calibration curve was obtained.

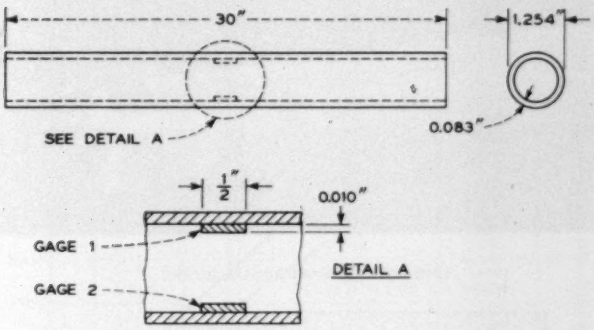


Fig. 1 — Steel tubing, showing location of strain gages.

The experimental arrangement for calibration of gages is shown in Fig. 2.

After calibration of gages, the polyethylene outer cylinder, whose inner diameter was 0.004 inch smaller than the outer diameter of the steel cylinder, was forced onto the steel cylinder in a universal testing machine. The assembled model appeared as in Fig. 3. Before the model was subjected to external hydrostatic pressure, the strain gages were balanced to indicate zero strain by adjusting resistances in a multichannel switching-and-balancing unit. The experimental arrangement for the model is given in Fig. 4. It is not the intent of the present paper to evaluate the pressure produced at the interface by the interference fit. External hydrostatic pressure was applied in increments of 500 psi up to 2500 psi and corresponding strains were observed. The pressure was reduced to zero and the reproducibility of the data was checked.

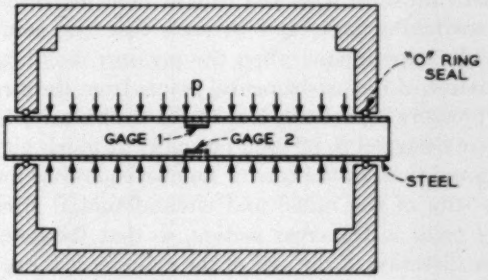


Fig. 2 — Experimental arrangement for calibration.

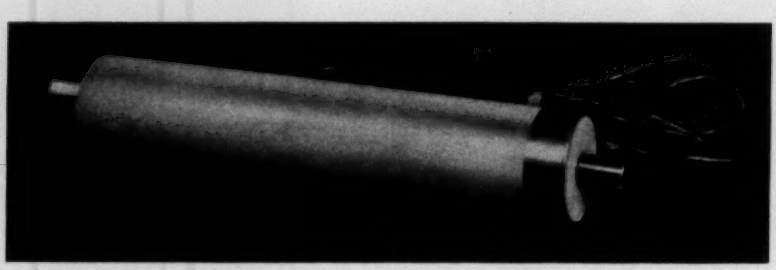


Fig. 3 — Experimental model.

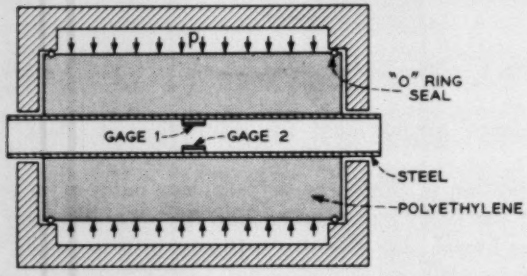


Fig. 4 — Experimental arrangement for the model.

III. EXPERIMENTAL RESULTS

The calibration curves for external hydrostatic pressure versus the strain observed from gages 1 and 2, when the external hydrostatic pressure was applied directly on the pipe, are given in Figs. 5 and 6. Corresponding curves when the external pressure was applied on the model are also plotted in these figures.

The lines pass through the origin because the bridge was initially balanced at zero pressure with the help of blancing resistances in the multichannel switching unit, and in each case the zero reading was checked after the experiment, when the pressure was brought back to zero. The deviation of the experimental points from the straight lines for low values of pressure is probably due to the nonlinearity in the pressure gages, which are designed to be used normally at much higher pressures.

It may be noted that, in the case of Lame's equations for thick-walled cylinders, the sum of the radial and circumferential stresses is a constant at every point in the cross section, so that the deformation of all elements in the direction of the axis of the cylinder is the same and cross sections of the cylinder remain plane after deformation. Hence, in spite of the considerable length of the cylinder, the experimental arrangement gives the verification for plane stress conditions.

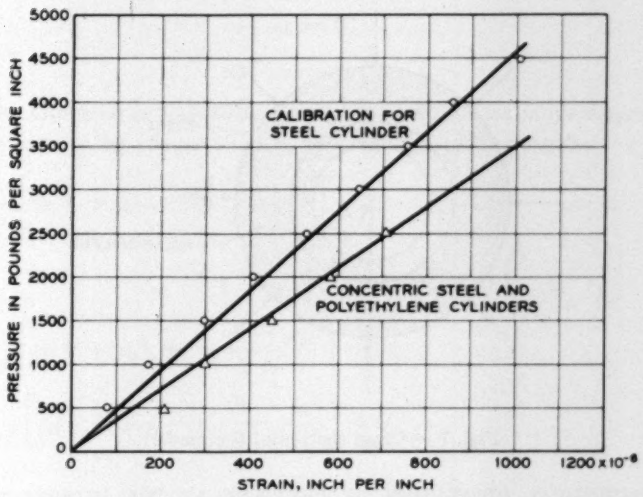


Fig. 5 — Pressure vs. strain from gage 1.

The pressure at the interface is obtained by drawing a vertical line from the observed value of strain to the calibration line for the steel cylinder alone and reading the corresponding pressure. Numerical calculations from Figs. 5 and 6 show that the pressure at the interface, in the case of plane stress is 1.3 times the external pressure.

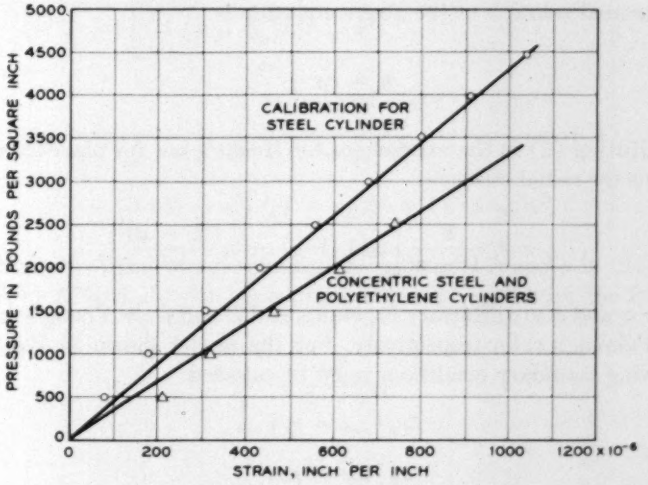


Fig. 6 — Pressure vs. strain from gage 2.

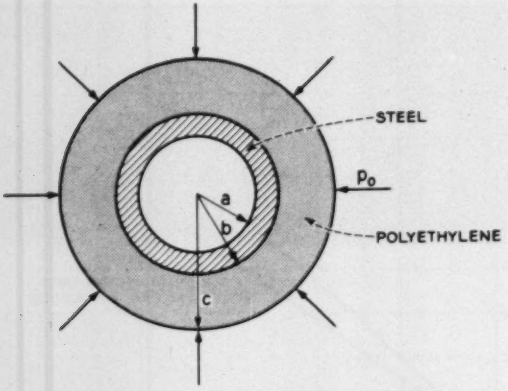


Fig. 7 — Cross section of the model.

IV. THEORETICAL DERIVATION OF PRESSURE AT THE INTERFACE

The differential equation of equilibrium,[1] in terms of radial displacement, for a thick-walled cylinder subjected to external and internal pressures is

$$\frac{d^2u}{dr^2} + \frac{1}{r}\frac{du}{dr} - \frac{u}{r^2} = 0. \tag{1}$$

The general solution of the above equation is

$$u = c_1 r + \frac{c_2}{r}. \tag{2}$$

Substituting (2) in the expressions for Hooke's law for plane stress, one obtains for radial stress, σ,

$$\sigma = \frac{E}{1-\mu^2}\left[c_1(1+\mu) - c_2\frac{(1-\mu)}{r^2}\right], \tag{3}$$

where c_1 and c_2 are arbitrary constants and E and μ are Young's modulus and Poisson's ratio respectively. For the model shown in Fig. 7, the following boundary conditions must be satisfied:

$$\sigma_p \mid_{r=b} = \sigma_s \mid_{r=b}, \tag{4}$$

$$\sigma_p \mid_{r=c} = p_0, \tag{5}$$

$$\sigma_s \mid_{r=a} = 0, \tag{6}$$

$$u_p \mid_{r=b} = u_s \mid_{r=b}. \tag{7}$$

In the above equations subscripts p and s stand for polyethylene and steel. Substituting (2) and (3) in (4) through (7), one gets:

$$\frac{E_p}{1-\mu_p^2}\left[c_1(1+\mu_p) - c_2\left(\frac{1-\mu_p}{b^2}\right)\right] = \frac{E_s}{1-\mu_s^2}\left[c_3(1+\mu_s) - c_4\left(\frac{1-\mu_s}{b^2}\right)\right], \tag{8}$$

$$\frac{E_p}{1-\mu_p^2}\left[c_1(1+\mu_p) - c_2\frac{(1-\mu_p)}{c_2}\right] = p_0, \tag{9}$$

$$\frac{E_s}{1-\mu_s^2}\left[c_3(1+\mu_s) - c_4\frac{(1-\mu_s)}{a^2}\right] = 0, \tag{10}$$

$$c_1 b + \frac{c_2}{b} = c_3 b + \frac{c_4}{b}. \tag{11}$$

The pressure p at the interface is obtained by solving for the constants from the foregoing equations and evaluating

$$\sigma_p \mid_{r=b} = p.$$

Equation (12) gives the pressure at the interface for a plane stress solution.

$$p = p_0\left\{\frac{(b^2-c^2)\left[\frac{E_s}{E_p}(1-\mu_p)(b^2-a^2) - (1-\mu_s)b^2 - (1+\mu_s)a^2\right]}{(b^2-c^2)[(1-\mu_s)b^2+(1+\mu_s)a^2] - \frac{E_s}{E_p}(b^2-a^2)[(1-\mu_p)b^2+(1+\mu_p)c^2]} + 1\right\}. \tag{12}$$

To obtain the plane strain solution, the constants E and μ in (12) are replaced by E^* and μ^*, with appropriate subscripts given by the following:[2]

$$E^* = \frac{E}{1-\mu^2}, \tag{13}$$

$$\mu^* = \frac{\mu}{1-\mu}. \tag{14}$$

V. NUMERICAL EXAMPLE FOR THE MODEL

The values for the various constants are:

$$a = 1.088 \text{ inches},$$

$$b = 1.254 \text{ inches},$$

$$c = 4.621 \text{ inches},$$

$$E_s = 29 \times 10^6 \text{ psi},$$

$$E_p = 19 \times 10^3 \text{ psi},$$

$$\mu_s = 0.3,$$

$$\mu_p = 0.5.$$

Substitution of above values in (12) gives for the plane stress solution

$$p = 1.3(p_0). \tag{15}$$

Replacing E and μ by (13) and (14) and substituting in (12), one obtains for the plane strain solution

$$p = 1.0(p_0). \tag{16}$$

It may be noticed from (12) that the pressure at the interface appears to depend on the values of E_p and μ_p. However, examination reveals that (12) is relatively insensitive to E_p, for the range of known values for low and intermediate density polyethylene. Since it is difficult to determine μ_p, one may use the previously known values of bulk modulus, K_p, for polyethylene and obtain μ_p from

$$K_p = \frac{E_p}{3(1 - 2\mu_p)}. \tag{17}$$

Refs. 3 and 4 give the value of $K_p \approx 0.5 \times 10^6$ psi. Ref. 5 gives for E_p, for low and intermediate density, the range of values from 19×10^3 to 55×10^3 psi. With $K_p = 0.5 \times 10^6$ psi and $E_p = 19 \times 10^3$ or 55×10^3 psi, (17) gives μ_p to be very nearly equal to 0.5. It may be noted that the analysis for concentric cylinders given in this paper agrees with experimental results when μ_p is very nearly equal to 0.5.

VI. CONCLUSION

Experiments indicate that an analysis based on equations of the two-dimensional theory of elasticity for the pressure at the interface of polyethylene outer and steel inner concentric cylinders, under plane

stress conditions subjected to a uniform pressure on the curved surface, is valid even for polyethylene, which is a material with nonlinear stress-strain relationship. The agreement between the theoretically calculated and experimentally determined values is very good. It is also shown that the value of Poisson's ratio for low and intermediate density polyethylene, for an elastic analysis, should be very nearly 0.5. The experimental verification of the plane stress solution given in the present paper confirms this value.

An experimental verification of the plane strain solution involves a number of difficulties in satisfying the condition that the axial strain be zero; it has not been attempted in the present paper. However, one may assume that the plane strain solution is also valid, because the plane stress solution obtained by using the same two-dimensional theory of elasticity has been experimentally verified. The plane strain solution gives for the numerical example the result that the pressure at the interface is the same as external pressure.

The cable is actually under a uniform pressure in three dimensions. The two-dimensional plane strain solution may be extended to a three-dimensional case of uniform pressure by a superposition of an axial loading of intensity p_0 and satisfying the appropriate boundary and compatibility conditions. However, since the value of Poisson's ratio for polyethylene is 0.5 and the area of polyethylene is much greater in comparison to the area of the steel, the pressure at the interface of the inner conductor and the polyethylene dielectric, as given by the plane strain solution, would not be much altered. Hence, the pressure at the interface of the inner conductor and the dielectric of the armorless ocean cable is equal to or at most slightly greater than the external pressure.

It may be remarked that a new experimental method employing resistance wire strain gages has been developed to measure interface pressures accurately. The method devised in this paper could be used in other applications to measure interface pressures.

VII. LIST OF SYMBOLS

a = inner radius of steel cylinder (inches).
b = outer radius of steel cylinder (inches).
c = outer radius of polyethylene cylinder (inches).
K_p = bulk modulus for polyethylene (psi).
p = pressure at the interface (psi).
p_0 = external hydrostatic pressure (psi).
u = radial displacement (inches).
r = radial distance (inches).

E = Young's modulus (psi).
E_p = Young's modulus for polyethylene (psi).
E_s = Young's modulus for steel (psi).
μ = Poisson's ratio.
μ_p = Poisson's ratio for polyethylene.
μ_s = Poisson's ratio for steel.
σ = radial stress (psi).
σ_p = radial stress in polyethylene (psi).
σ_s = radial stress in steel (psi).

REFERENCES

1. Timoshenko, S., *Strength of Materials*, Part II, D. Van Nostrand Co., New York, 1956, pp. 205–214.
2. Timoshenko, S. and Goodier, J. N., *Theory of Elasticity*, McGraw-Hill Book Co., New York, 1951, pp. 34; 54–60.
3. Bridgman, P. W., Proc. Amer. Acad. Arts & Sci., **76**, 1948, p. 71.
4. Weir, C. E., National Bureau of Standards, Research Paper 2192, 1951.
5. *Modern Plastics Encyclopedia*, Plastics Properties Chart, Breskin Publications, New York, 1959.

The Construction of Missile Guidance Codes Resistant to Random Interference

By A. R. ECKLER

(Manuscript received February 8, 1960)

Many types of missiles are guided by a finite set of distinct commands radioed from the ground in the form of a time-sequence of RF *pulses. The command information is contained in the $n - 1$ time spacings between successive pulses in a group of size n, and is encoded and decoded by means of multitapped delay lines combined with* AND *gates. This paper discusses the problem of encoding command information (i.e., selecting the time spacings between pulses) so that m false pulses ($m \leqq n - 2$) cannot combine with the n true pulses in any way to form a false command. Although it is very easy to state the restrictions that must be imposed on the time spacings between the pulses in the different commands, no general methods exist for finding, among codes satisfying these restrictions, those codes in which the longest command is as short as possible. This paper presents certain lower bounds, together with a few empirically derived codes approaching these lower bounds. The relationship between these codes and the well-known error-correcting binary codes of information theory is discussed in an appendix.*

I. INTRODUCTION

Many types of missiles are guided by commands consisting of a time sequence of RF pulses. The missile receiver is capable of receiving not only the true command pulses, but also any interfering RF pulses emanating from other radars in the vicinity (friendly interference) or deliberately generated by the enemy (enemy interference). The purpose of this paper is to suggest command-encoding methods that minimize the effects of this interference. Greatest emphasis is placed on random interference models, because it is difficult to make realistic general assumptions about the behavior of an enemy who knows something about the code structure.

The following sections may convey the impression that the construc-

tion of a code invulnerable to random pulse interference is the only factor entering into the choice of a code. This is not meant to be the case; accuracy and reliability of decoding equipment, weight in the missile receiver, power considerations at the transmitter and missile dynamics and maneuverability also play very important roles. However, it will be made clear to the reader that code invulnerability should not be completely ignored, and that reasonable guideposts of design are available.

In order to fix ideas, the codes presented in this paper are discussed in terms of missile guidance. However, the reader will note that other applications are possible. In particular, the codes described here can be used in radio communications to provide protection against the mutual interference of many communications transmitters operating in the same area on the same frequency.

1.1 *The Engineering Background of the Problem*

In many missile guidance systems, command information is sent to the missile by means of RF pulses. Specifically, the information is contained in the time spacings between successive pulses in a group of two or more pulses. Successive groups of pulses are spaced much more widely apart because of average power limitations on the magnetron.

There are many physical devices suitable for encoding and decoding commands. This paper discusses a typical device for encoding discrete commands, such as "yaw left one unit" or "pitch down one unit." These commands are added up in the missile receiver until the desired correction has been achieved. Discrete commands can also be used for such functions as turning rocket motors on or off or destroying the missile.

A discrete command is conveniently decoded in the missile receiver by means of a lumped-parameter or distributed-parameter delay line tapped at several points (see Fig. 1). The earlier pulses in an incoming sequence are delayed the correct amounts by the delay line in order that the inputs to the AND gate be simultaneously activated.

In practice, it is sufficient that the several AND gate inputs be activated within a short time interval. Let τ denote the maximum possible range of input times (i.e., the difference in time between the earliest and latest input) that still activates the AND gate. Because of τ (which is called the discrete command tolerance), one must be careful not to let two distinct commands have similar encodings. It is shown in the Appendix that an adequate separation of different commands is always achieved if one assigns time spacings in integral multiples of the discrete command tolerance τ.

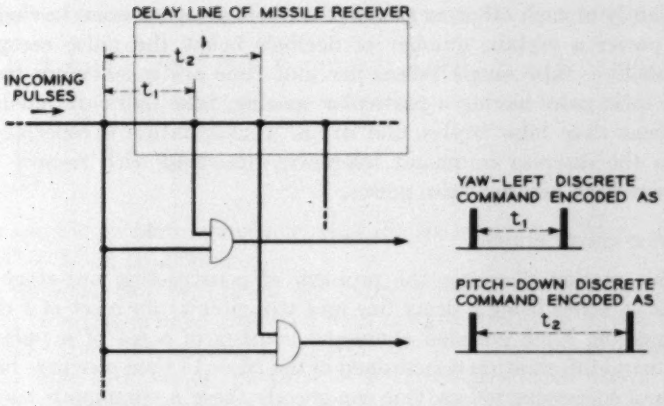


Fig. 1 — Method of decoding discrete command by means of distributed-parameter delay line tapped at several points.

It is clear that only two pulses are needed to encode a command. However, in order to obtain a greater degree of security against false pulses, it is necessary to use three or more pulses per command. The mathematical problem discussed in this paper is the construction of such security codes.

The same delay line and AND gate can be used in the communications problem already mentioned. Here the group of pulses plays the role of an address code rather than a command. Each communications transmitter uses a specific address code as a generalized "pulse" with which to communicate information; thus, different transmitters can use the same frequency without interfering with each other.

1.2 *A Criterion for Judging Code Effectiveness*

The invulnerability of codes to false pulses is measured by the following simple criterion: the *maximum* number of false pulses that cannot interfere with a code group (either by changing the position of a command or producing a different command), no matter how these false pulses are placed in time with respect to the true code group.

This criterion permits only a rough ordering of codes with respect to invulnerability; redundancy of orders and closed-loop missile response also play an obvious role. However, it does focus attention on a very important factor. No matter what random interference model is postulated — for example, a Poisson process of random pulses, a set of radars with approximately the same pulse repetition frequency operating inde-

pendently of each other, or peaks of random gaussian noise having average power a certain number of decibels below the pulse recognition threshold — false single pulses per unit time are much more frequent than false pairs having a particular spacing, false pairs are much more frequent than false triples, and so on. This situation is especially true when the discrete command tolerance τ is small with respect to the average time between false pulses.

II. ONE-STAGE CODES

This section discusses the problem of constructing one-stage codes (that is, codes using a delay line and AND circuit) for a set of k distinct commands. Each encoded command consists of a set of n pulses; the command information is contained in the $(n - 1)$ time spacings between pairs of successive pulses. One can encode these k commands such that i false pulses cannot combine in any way with $(n - i)$ pulses from any command to form either the same command (shifted in time) or one of the other $(k - 1)$ commands. The maximum possible value of i is equal to $(n - 2)$, because $(n - 1)$ false pulses can combine with one true pulse to form false commands in many different ways.

We first discuss methods of encoding commands invulnerable to $(n - 2)$ false pulses; later this restriction is relaxed to fewer pulses.

2.1 *The Construction of One-Stage Codes with Maximum Protection*

It may be helpful to think about a specific coding problem while reading this section. Let us assume that we wish to encode two commands $(k = 2)$ consisting of three pulses each $(n = 3)$; by suitable encoding we can insure that a single false pulse $(n - 2 = 1)$ cannot combine with two pulses to form any false command. This example will be referred to below by sentences enclosed in brackets.

As mentioned earlier, it is convenient to make all the time spacings of a discrete code integral multiples of the discrete command tolerance τ, in order to avoid overlapping problems between different commands. This being so, one can represent the various time spacings with the integers $1, 2, 3, \cdots$; the jth n-pulse command is represented by an $(n - 1)$-dimensional vector of integers $(t_1^j, t_2^j, \cdots, t_{n-1}^j)$. [Specifically, the first command is given by the pair of integers (t_1^1, t_2^1) and the second command by (t_1^2, t_2^2)].

The one-stage coding problem for discrete commands can now be stated in mathematical terms. Assume that one has k distinct n-pulse commands. All these commands are effective against $(n - 2)$ false pulses [in the sense discussed above] if and only if the following $kn(n - 1)/2$ integers are *all* different:

$$\begin{cases} t_i^j & \text{for } i = 1, 2, \cdots, n-1 \text{ and } j = 1, 2, \cdots, k; \\ t_i^j + t_{i+1}^j & \text{for } i = 1, 2, \cdots, n-2 \text{ and } j = 1, 2, \cdots, k; \\ \vdots & \\ \sum_{i=1}^{n-1} t_i^j & \text{for } j = 1, 2, \cdots, k. \end{cases}$$

[In our specific problem, we require that the six integers $t_1^1, t_2^1, t_1^1 + t_2^1, t_1^2, t_2^2, t_1^2 + t_2^2$ all be different.]

The theorem can be proved by indirect reasoning. Suppose that $(n - 2)$ false pulses have combined with two true pulses to form a false command. Then the spacing t_0 between the two true pulses must be equal to *more than one* of the $kn(n - 1)/2$ integers listed above (the set of all possible spacings between pairs of pulses in the k commands). For, if t_0 were equal to only one of these integers, the only command that could be formed with the aid of these two true pulses would be the correct command. Conversely, if two of these integers are the same, it is easy to construct a false command using $(n - 2)$ properly spaced false pulses. [For example, if the two commands are (3, 8) and (2, 6), then the second command can be falsely formed from the first command by placing an extra pulse two time units after the second pulse in the first command. In this case, the six integers are 3, 8, 11, 2, 6, 8, and t_0 is equal to 8.]

If one allows the integers t_i^j to be arbitrarily large, there is no difficulty in discovering suitable codes. [For example, (31, 73) and (9, 45).] However, equipment and time restrictions usually make it desirable to keep all the commands as short as possible (measured from first to last pulse). Therefore, one may pose the following interesting mathematical problems:

i. For a given k and n, what is the minimum possible length of the longest command under the restrictions above?

ii. Are there any (simple and practical) methods for constructing codes with minimum command length (or a little longer)?

Neither of the above two problems has been solved; lower bounds can (in a limited way) be set for the first, and the second has largely been investigated by trial and error. The problem is similar to the eight queens problem* in chess, but with additional restrictions. More gener-

* The eight queens problem consists of placing eight queens on the chess board so that no queen can capture any other along a row, column, or diagonal. A specific code can be characterized by a pair of integers (a, b); this denotes a position on a generalized chessboard (not restricted to the eight-by-eight size).

ally, it appears to be a problem in partitions in the theory of numbers.

A general lower bound L for the minimum possible length of the longest command is given by $kn(n-1)/2$. However, this bound cannot always be achieved even for three-pulse codes. For a three-pulse code, it is true that

$$\sum_{i=1}^{2}\sum_{j=1}^{k} t_i^j = \sum_{j=1}^{k} (t_1^j + t_2^j).$$

If the longest command length is equal to $3k$, the set of integers

$$t_i^j (i = 1, 2; j = 1, 2, \cdots, k),\ t_1^j + t_2^j (j = 1, 2, \cdots, k)$$

is a rearrangement of the set of integers $(1, 2, \cdots, 3k)$. Therefore, the sum of the two sides of the above equation is $3k(3k+1)/2$, and the sum of one side alone is $3k(3k+1)/4$. But this last quantity is an integer if and only if $k \equiv 0 \pmod 4$ or $k \equiv 1 \pmod 4$. In other words, $3k$ is a lower bound L when $k = 1, 4, 5, 8, 9, 12, 13, \cdots$, and $3k+1$ is a lower bound L for all other k.

2.2 *Some Typical One-Stage Encodings*

We now list specific discrete encodings for several values of k and n. For the three-pulse codes the lower bound L for the longest delay line needed has actually been achieved for $k = 1\ (1)\ 10$. The first number gives the time spacing between the first and second pulses, and the second number the time spacing between the second and third pulses:

k	L	Typical minimum-length discrete-command three-pulse encodings									
2	7	(1,5)	(3,4)								
3	10	(4,6)	(1,7)	(2,3)							
4	12	(1,9)	(3,8)	(5,7)	(2,4)						
5	15	(3,12)	(5,9)	(2,11)	(4,6)	(1,7)					
6	19	(4,15)	(2,16)	(5,12)	(3,8)	(1,9)	(6,7)				
7	22	(3,19)	(5,16)	(7,13)	(6,12)	(2,15)	(4,10)	(1,8)			
8	24	(1,23)	(2,11)	(3,18)	(4,10)	(5,15)	(6,16)	(7,12)	(8,9)		
9	27	(1,26)	(3,22)	(9,15)	(2,21)	(6,14)	(7,12)	(10,8)	(4,13)	(5,11)	
10	31	(5,26)	(14,16)	(2,27)	(8,20)	(4,21)	(9,15)	(1,22)	(7,12)	(3,10)	(6,11)

Undoubtedly this list could be indefinitely extended, since it is easy to discover many different encodings for any k which have the same length for a maximum-length command.

Unfortunately, an analogous method for improving the lower bound L does not exist for $n > 3$. As n increases, it becomes more difficult to keep the minimum delay line small, and $kn(n-1)/2$ becomes a very unrealistic lower bound.

The tables below list representative discrete-command four-, five- and six-pulse encodings; the reader is invited to find an encoding with a smaller L' if he can (L' denotes the minimum command length that has been achieved, but it may not be the true minimum):

k	L	L'	Typical discrete-command four-pulse encodings				
2	12	13	(2,5,6)	(3,1,8)			
3	18	21	(1,5,12)	(2,8,11)	(3,4,9)		
4	24	27	(1,11,15)	(2,6,17)	(3,7,14)	(4,5,13)	
5	30	31	(2,20,9)	(3,14,13)	(4,8,16)	(5,10,11)	(6,1,18)

k	L	L'	Typical discrete-command five-pulse encodings		
2	20	25	(1,8,3,13)	(2,5,10,4)	
3	30	38	(1,5,12,20)	(2,8,11,14)	(3,4,9,15)

k	L	L'	Typical discrete-command six-pulse encodings	
2	30	45	(1,8,3,13,20)	(2,5,10,4,18)

If encodings not given in these tables are desired, it should be possible to use a high-speed digital computer to search through a large number of encodings (either systematically or at random) and print out the minimum-length encoding that it finds. Unless an algorithm for computing encodings is found, this is the only practical method available.

2.3 *A General Class of Discrete Codes*

As noted in Section 2.2, the length of the longest command increases rapidly with n, the number of pulses in the command. This, of course, is the price that one pays for protecting oneself against $(n - 2)$ false pulses arranged in any pattern whatever. One can always reduce the code length by reducing the protection; that is, construct minimum-length n-pulse codes that protect against $(n - 1)$ or fewer pulses ($i = 3, 4, 5, \cdots, n - 1$). The following formulation of the problem is more realistic. Suppose that one wishes protection against m false pulses combining with $(n - m)$ pulses in any command.

i. How does one construct n-pulse codes ($n > m + 2$), keeping the length L of the longest command as short as possible?

ii. If power is no limitation (any n can be used within reason), for what value of n is L minimized?

No general solution to either of these problems is known. This section presents typical encodings for a few of the simplest codes, which were found by trial-and-error methods.

The simplest problem in this class is the construction of a four-pulse

code invulnerable to a single false pulse. It can be proved (by the method used in Section 2.2) that, in order for a four-pulse, k-command code to be invulnerable to one false pulse the following $4k$ integer-pairs must all be different:

$$(t_1^j, t_2^j), \quad (t_2^j, t_3^j), \quad (t_1^j + t_2^j, t_3^j), \quad (t_1^j, t_2^j + t_3^j)$$

$$\text{for } j = 1, 2, \cdots, k.$$

As before, a lower bound L for the length of the longest command can be obtained. There is one integer-pair (1,1) which sums to 2; there are two more which sum to 3, and in general there are $n(n-1)/2$ with sums less than or equal to n. Therefore, a lower bound L is given by the smallest value of n satisfying $4k \leqq n(n-1)/2$. Most of the specific discrete encodings listed below actually achieve this bound; the ones that exceed it by one are marked with an asterisk:

k	L	Typical four-pulse encodings invulnerable to one false pulse						
1	4	(1,1,2)						
2	5	(1,3,1)	(2,1,2)					
3	6	(4,1,1)	(1,2,2)	(2,1,3)				
4	7	(1,4,2)	(2,4,1)	(1,1,2)	(3,2,1)			
5	7	(1,3,1)	(3,2,3)	(4,2,1)	(1,2,4)	(2,5,1)*		
6	8	(1,5,2)	(2,5,1)	(1,2,4)	(4,2,1)	(3,2,3)	(1,3,1)	
7	8	(1,5,3)	(2,2,5)	(3,3,2)	(4,2,1)	(7,1,1)	(1,3,4)	(2,3,1)*

For moderate values of k, the minimum length has been reduced more than 50 per cent by adding one more pulse to the code.

In order to construct a five-pulse k-command code that is invulnerable to two false pulses, it can be shown (by the method used in Section 2.2) that the following $10k$ integer-pairs must all be different:

$$(t_1^j, t_2^j), \quad (t_2^j, t_3^j), \quad (t_3^j, t_4^j),$$

$$(t_1^j + t_2^j, t_3^j), \quad (t_2^j + t_3^j, t_4^j), \quad (t_1^j, t_2^j + t_3^j), \quad (t_2^j, t_3^j + t_4^j),$$

$$(t_1^j, t_2^j + t_3^j + t_4^j), \quad (t_1^j + t_2^j, t_3^j + t_4^j), \quad (t_1^j + t_2^j + t_3^j, t_4^j)$$

$$\text{for } j = 1, 2, \cdots, k.$$

The lower bound L is given by the smallest value of n satisfying $10k \leqq n(n-1)/2$. Three typical encodings are given below; it it believed that they are the shortest ones possible:

k	L	Typical five-pulse encodings invulnerable to two false pulses		
1	5	(1,2,2,1)		
2	7	(2,3,1,2)	(1,3,4,1)	
3	9	(3,3,1,4)	(2,2,1,6)	(1,7,1,2)

In order to construct a five-pulse k-command code that is invulnerable to one false pulse, it can be shown that the following $5k$ integer-triples must all be different:

$$(t_1^j, t_2^j, t_3^j), \quad (t_2^j, t_3^j, t_4^j),$$

$$(t_1^j + t_2^j, t_3^j, t_4^j), \quad (t_1^j, t_2^j + t_3^j, t_4^j), \quad (t_1^j, t_2^j, t_3^j + t_4^j)$$

$$\text{for } j = 1, 2, \cdots, k.$$

The lower bound L is given by the smallest value of n satisfying $5k \leqq n(n-1)(n-2)/6$. A few typical encodings (believed to be minimum length) are given below:

k	L	Typical five-pulse encodings invulnerable to one false pulse			
1	5	(1,1,2,1)			
2	5	(1,1,3,1)	(2,1,1,2)		
3	6	(1,1,3,1)	(2,1,1,2)	(1,3,2,1)	
4	6	(1,1,3,1)	(2,1,1,2)	(1,3,2,1)	(2,2,1,2)

As mentioned before, these tables can be extended with the aid of a high-speed digital computer.

The following table summarizes the known behavior of one-stage discrete-command codes. It records the shortest known length of the longest command; most of these lengths are believed to be the smallest ones achievable. It is evident that one can often reduce this length quite markedly by adding pulses to the code, and that the savings are likely to be greater as the number of commands increases:

Codes Invulnerable to One False Pulse

	Number of pulses		
k	3	4	5
1	3	4	5
2	7	5	6
3	10	6	7
4	12	7	7
5	15	8	—
6	19	8	—

Codes Invulnerable to Two False Pulses

	Number of pulses		
k	4	5	6
1	7	6	—
2	13	9	—
3	21	11	—
4	27	—	—

III. TWO-STAGE CODES

The single-stage codes discussed in Section II should prove quite useful to the designer of missile guidance equipment. However, they suffer from the drawback of trying to do too many things at once. Sometimes it is more reasonable to break the decoding job up into several smaller jobs that are performed in sequence. Thus, the output pulse resulting

from one decoding operation becomes an input pulse to the next decoding operation.

An important advantage of such a break-up is that at least part of the security function of the code can be separated from the command function. This means that one can build in between-missile security without being forced to change the command codes from one missile to the next. Furthermore, the between-missile security codes can be used to increase the search time of a knowledgeable enemy for the correct code.

A second advantage is greater simplicity of both codes and components. If one needs to protect against (say) as many as four false pulses, it is not easy to find optimum single-stage codes, and the required delay lines may not be available. Multistage codes provide additional designs that may be easier to instrument. Balanced against this, of course, may be an increase in the number of components needed.

Multistage discrete-command codes are those that use a delay line and multiple AND circuit decoding mechanism (see Section I) at each stage. For simplicity, let us restrict ourselves to two-stage codes, which possess most of the potential advantages of multistage codes. Let the symbol $i \mid j$ denote a two-stage code containing ij pulses: the first stage decodes j clusters of i pulses each and the second stage decodes the cluster of j output pulses emitted by the first-stage AND circuit.

Usually the command part of the code occurs at the second stage. However, it makes no difference whatever which stage it is assigned to: an $i \mid j$ security-command code has precisely the same invulnerability to false pulses as does a $j \mid i$ command-security code. Let us assume for convenience that the commands are always contained in the *second* stage.

3.1 *An Upper Bound for the Protection*

How much protection can an $i \mid j$ code give against false pulses? Assume that the two stages are each designed according to single-stage rules, so that they are invulnerable to $(i - 2)$ and $(j - 2)$ false pulses, respectively. Then an upper bound to the number of false pulses that can be arranged in any pattern whatever with the ij true pulses *without* forming a false (or repeated) command is

$$M = \min [i(j - 1) - 1, j(i - 1) - 1].$$

The first term arises because false pulses can be arranged in $(j - 1)$ groups of i false pulses each, and these groups can be combined with one true i-pulse group to form a false command. The second term arises

Fig. 2 — First- and second-term error for 2 | 3 code.

because false pulses can be arranged in j groups of $(i - 1)$ false pulses each, augmented by j true pulses, one from each of the j groups in the correct code. This is illustrated in Fig. 2 for the 2 | 3 code.

Although M is the same for both an $i \mid j$ and a $j \mid i$ code, the two errors are not symmetric. Note that the first type of error can result in a false command, but the second type of error can only repeat the *same* command slightly earlier or later in time. The second error is likely to be less serious, and in fact can be eliminated by introducing a device that prevents repetition of the same command within a specified period of time. Accordingly, in the rest of this section we emphasize codes that are protected against $i(j - 1) - 1$ false pulses combining with true pulses to form a *different* false command.

For i, j greater than or equal to two, the maximum possible protection is always less than that achievable with a single-stage code using ij pulses. Thus, these codes are somewhat comparable to the reduced-protection single-stage codes discussed in Section II.

3.2 *Methods for Combining Security and Command Codes*

In order to actually achieve the upper bound of the protection, one must be a little careful when combining the two stages of the code. There are many ways of doing this, and the choice of a particular method depends upon the ease of instrumentation. For example, consider the following alternatives for the 2 | 3 code (see Fig. 3).

3.2.1 *Short Command Codes, Long Security Codes*

Let the three-pulse commands be encoded according to the methods in Section II. The longest command will be approximately $3k$ if there are k different commands. The two pulses forming the security code should be more than $6k$ apart in order to avoid the possibility of three false pulses forming a false command. This design is probably the easiest to instrument.

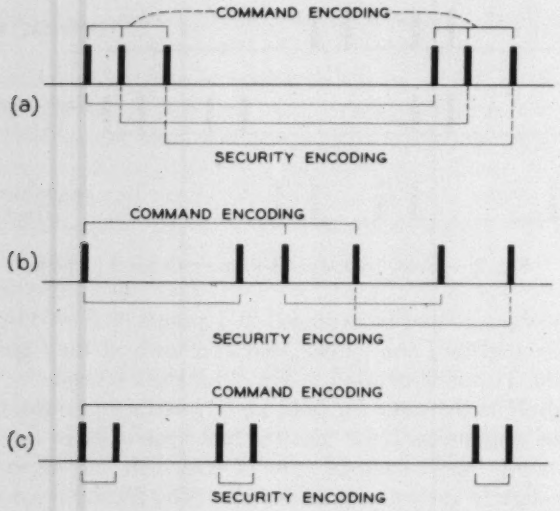


Fig. 3 — Methods for combining security and command codes: (a) short command codes, long security codes: (b) interleaved security and command codes; (c) long command codes, short security codes.

3.2.2 *Interleaved Security and Command Codes*

Let the spacing between the two-pulse security codes consist of *odd* integers, and encode the three-pulse commands in *even* integers using the methods discussed in Section II. To illustrate: if four commands are to be sent, let them be (2, 18), (6, 16), (10, 14) and (4, 8). More security and command codes can be easily added.

3.2.3 *Long Command Codes, Short Security Codes*

Let the spacing between the two-pulse security codes be given by the integers 1, 2, $\cdots$, m. Encode the three-pulse commands in multiples of $2m$ (or greater) according to the methods of Section II. This is likely to be the most difficult design to instrument.

These alternatives all provide the maximum possible protection; four false pulses are needed to form a false command, and three false pulses are needed to shift the true command in time.

The first and third alternatives are very easy to encode for any values of i and j (if the necessary single-stage codes are available), and are not discussed further. False pulses that combine with true pulses to form false

security codes cannot possibly form false commands as well, because these false pulses are either too far from or too close to the other true pulses to do the job. Frequently the longest delay line is somewhat longer for these alternatives than it would be if an interleaved code were used. Accordingly, we devote the rest of this part of the paper to the construction of interleaved codes.

3.3 *The Problem of Interleaving Security and Command Codes*

This section discusses the problem of interleaving i-pulse security codes with j-pulse command codes in such a way that one is protected against $i(j-1)-1$ false pulses forming a false command. Probably the simplest way to interleave the codes is to reserve the integers $f(i)$, $2f(i)$, $3f(i)$, $\cdots$ for the command codes. Then one can construct minimum-length command codes invulnerable to $(j-2)$ false input pulses according to the methods discussed in Section II. [It will be shown later that the minimum possible value for $f(i)$ is i.]

The interleaving problem is now reduced to the problem of selecting the security code spacings so that one is protected against $i(j-1)-1$ false pulses. The following sections present two restrictions on the security codes and show that these restrictions are sufficient to guarantee this protection. No claim is made that this is the only way to interleave codes; however, the resulting codes do have optimum properties, which are described later.

3.3.1 *Two Restrictions on the Security Code*

The security code spacings $(t_1, t_2, \cdots, t_{i-1})$ should be chosen in such a way that

(a) no set of $(i-2)$ false pulses can combine with two true pulses to form a false security code, and

(b) none of the $i(i-1)$ integers

$$\begin{cases} \pm t_k, & k = 1, 2, \cdots, i-1; \\ \pm(t_k + t_{k+1}), & k = 1, 2, \cdots, i-2; \\ \vdots \\ \pm(t_1 + t_2 + \cdots + t_{i-1}) \end{cases}$$

should be a multiple of $f(i)$ [that is, equal to 0 mod $f(i)$].

The first restriction states that security codes should be single-stage codes consisting of only one command (see Section II). The second restriction accomplishes two goals. First, since none of the integers t_1, .

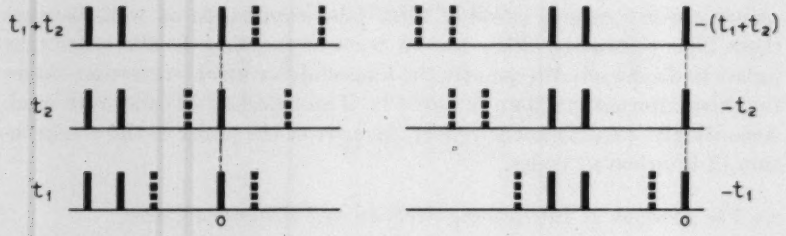


Fig. 4 — Six false security codes for a three-pulse code.

$t_1 + t_2, \cdots, t_1 + t_2 + \cdots + t_{i-1}$ is a multiple of $f(i)$, any $i \mid j$ code will consist of the full set of ij pulses. (Any pulse serving two functions at the same time is a natural candidate for a false pulse.) Second, the above set of integers lists the starting times of all the false security codes consisting of $(i - 1)$ false pulses and one true pulse (relative to the starting time of the true security code). For example, for a three-pulse security code, the six false security codes are shown in Fig. 4.

Now, consider what happens if one of the above spacings is a multiple of $f(i)$. One of the command code spacings is formed by the false security code corresponding to this spacing and one of the true security codes; the other $(j - 2)$ command spacings are formed by $i(j - 2)$ false pulses. But $(i - 1) + i(j - 2) = i(j - 1) - 1$, which is the number of false pulses we wish to protect against. The $3 \mid 4$ code illustrates this point (see Fig. 5).

It is shown in Section 3.3.2 that codes satisfying these restrictions are protected against $i(j - 1) - 1$ false pulses, and that the minimum possible value of $f(i)$ is i.

3.3.2 *Proof That These Restrictions Insure Protection*

Given these two restrictions, how can false commands be formed? Each security code in the false command can include at most one true

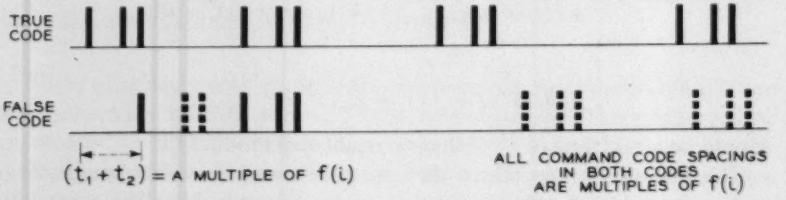


Fig. 5 — A $3 \mid 4$ code with spacing a multiple of $f(i)$.

pulse, because of restrictions (a) and (b). [There is one exception: a false command can be constructed out of one true security code and $(j - 1)$ false security codes of i pulses each.] Furthermore, ij pulses are needed for the false command, according to restriction (b). Therefore, if we can show that the false command contains at most i true pulses (one in each of i different false security codes), then we are done, for the $(ij - i)$ or more false pulses needed to make up the balance of the false command will exceed $i(j - 1) - 1$, the number of false pulses we wish to protect against. Since all the true command spacings are multiples of $f(i)$, it is sufficient to show that one can select at most i false security codes consisting of one true pulse and $(i - 1)$ false pulses each, which are all spaced $f(i)$ apart from each other.

In the preceding section, we listed all the $i(i - 1)$ possible false security code starting times (relative to the true security code). If we can show that exactly i of these times are equal to $a \pmod{i}$, where $a = 1, 2, \cdots, i - 1$, then we will have proved that at most i different false security codes can have the proper command spacing. Furthermore, we will have proved that $f(i) = i$; and it is clear that this is the smallest integer it can be.

Consider the following arrangement of the $i(i - 1)$ false security code starting times in a matrix with i rows and $(i - 1)$ columns:

$$\left\{\begin{array}{l} t_1, t_1 + t_2, t_1 + t_2 + t_3, \cdots \quad \cdots, t_1 + \cdots + t_{i-1} \\ t_2, t_2 + t_3, \cdots \quad \cdots, t_2 + \cdots + t_{i-1}, -t_1 \\ t_3, t_3 + t_4, \cdots \quad \cdots, t_3 + \cdots + t_{i-1}, -t_2, -(t_1 + t_2) \\ \quad \vdots \\ t_{i-1}, -t_{i-2}, -(t_{i-3} + t_{i-2}), \cdots \quad \cdots, -(t_1 + \cdots + t_{i-2}) \\ -t_{i-1}, -(t_{i-2} + t_{i-1}), \cdots \quad \cdots, -(t_2 + \cdots + t_{i-1}), \\ \qquad\qquad -(t_1 + \cdots + t_{i-1}). \end{array}\right.$$

If the integers in each row are reduced modulo i, they will be a permutation of the integers $1, 2, \cdots, i - 1$. We prove this by showing that the opposite conclusion leads to a contradiction. If two integers in a row are both equal to the same integer mod i, then their difference is equal to 0 mod i. But their difference is contained elsewhere in the matrix, which contradicts the assumption that no starting times are 0 mod i [restriction (b) above]. Since we have proved that each row is a permutation of the first $(i - 1)$ integers, it is clear that the matrix consists of i starting times equal to $a \pmod{i}$, $a = 1, 2, \cdots, (i - 1)$.

It is not too difficult to actually find security codes satisfying restriction (b). For example, consider the security code ($t_1 = 1, t_2 = 2, t_3 = 3$). We have

$$\begin{aligned}
\pm t_1 = \pm 1 &= 1 \bmod 4, 3 \bmod 4;\\
\pm t_2 = \pm 2 &= 2 \bmod 4, 2 \bmod 4;\\
\pm t_3 = \pm 3 &= 3 \bmod 4, 1 \bmod 4;\\
\pm (t_1 + t_2) = \pm 3 &= 3 \bmod 4, 1 \bmod 4;\\
\pm (t_2 + t_3) = \pm 5 &= 1 \bmod 4, 3 \bmod 4;\\
\pm (t_1 + t_2 + t_3) = \pm 6 &= 2 \bmod 4, 2 \bmod 4.
\end{aligned}$$

There are exactly four starting times for each of the three values of a. Note that this security code does not satisfy restriction (a), but that ($t_1 = 1, t_2 = 2 + 4 = 6, t_3 = 3$) does.

If $i < j$ in an $i \mid j$ two-stage code, the above two restrictions on the construction of security codes should be applied. However, if $i > j$, then we automatically have more than the minimum protection $i(j - 1) - 1$ against false pulses — the false command can contain at most $ij - j$ false pulses (ignoring the one exception mentioned above). One can trade off this unnecessary additional protection for a larger set of security codes. To be specific, one can replace the first restriction with one of the following less stringent restrictions:

(a′) no set of $(i - k)$ false pulses, for $k = 3, 4, \cdots, i - 1$, can combine with k true pulses to form a false security code.

These single-stage codes have already been discussed in Section II. For example, if one wishes to construct a $4 \mid 2$ code, then one can use security codes that protect against only one false pulse (instead of two, the greatest possible protection).

3.4 *Some Typical Interleaved Security and Command Codes*

The preceding section presented two restrictions on the security code, which guarantee that the resulting $i \mid j$ interleaved two-stage code will be protected against $i(j - 1) - 1$ false pulses forming a false command. Nothing, however, was said about the actual construction of such codes. This section presents sample codes for low values of i and j.

The command codes of an $i \mid j$ code are restricted to the spacings i, $2i$, $3i$, $\cdots$ and are constructed according to the methods discussed in Section II. The security codes are constructed by trial-and-error methods to satisfy the two restrictions. For $i = 2$, the odd integers all form se-

curity encodings. For $i = 3$, the pairs (1, 4), (2, 5), (4, 7), (5, 8), $\cdots$ and their reverses are legitimate security encodings. In general, such encodings must be of the form (t_1, t_2), where $t_1 \neq t_2$ and (t_1, t_2) equals (1 mod 3, 1 mod 3) or (2 mod 3, 2 mod 3). For $i = 4$, the triples (1, 5, 9), (2, 1, 6), (1, 6, 3), (2, 3, 6) and (6, 5, 2) are examples of short security encodings. In general, these security encodings must be of the form (t_1, t_2, t_3) where none of the integers $t_1, t_2, t_3, t_1 + t_2, t_2 + t_3$ are equal, and (t_1, t_2, t_3) is in one of the following forms: (1 mod 4, 1 mod 4, 1 mod 4); (2 mod 4, 1 mod 4, 2 mod 4); (1 mod 4, 2 mod 4, 3 mod 4); (2 mod 4, 3 mod 4, 2 mod 4); (3 mod 4, 3 mod 4, 3 mod 4) or (3 mod 4, 2 mod 4, 1 mod 4).

These interleaved two-stage codes can be regarded as minimum-length codes in the following sense: given that the command codes are restricted to the integers $f(i), 2f(i), 3f(i), \cdots$, then

(a) $f(i)$ is equal to its minimum value i;

(b) the longest command code is as short as possible; and

(c) the security codes are as short as possible.

Assuming that one wishes to protect against a given number of false pulses, which $i \mid j$ code should one select? The answer depends upon the relative importance of minimizing the number of pulses ij in the code, and keeping the longest command delay line as short as possible. There are two possible ways to balance these factors against each other:

(a) apply the methods used in one-stage codes (Section II) to the command code;

(b) change the values of both i and j; codes with large i and small j have shorter delay lines and more pulses, while codes with small i and large j have the opposite characteristics.

Let us illustrate these ideas with a few simple examples. Assume that we wish to transmit six commands. The $2 \mid 3$ and $4 \mid 2$ codes both provide protection against at most three false pulses. The number of pulses per command is six and eight, while the maximum delay-line length is $19(2) = 38$ and $6(4) = 24$, respectively. Suppose that one uses a $2 \mid 4$ code but designs the command code so that it is invulnerable to one false input pulse (no matter how placed) but not two. Then, referring to Section II, we see that the number of pulses per command is eight and the maximum delay line is $8(2) = 16$.

How do two-stage codes compare with one-stage codes? The only direct comparison available is the $2 \mid 2$ code with the four-pulse one-stage code protected against one false pulse: both codes have the same number of pulses and the same protection against false pulses. The length of the longest delay line needed is tabled below for various commands:

Number of commands	2	3	4	5	6	7
Two-stage code	4	6	8	10	12	14
One-stage code	5	6	7	8	8	9

IV. ACKNOWLEDGMENT

The author is indebted to W. J. Albersheim for the suggestion that a general study of missile codes be made. He is also indebted to W. L. Roach for a very careful reading of an earlier draft; his suggestions led to many improvements in this paper.

APPENDIX

Relationship to Error-Correcting Binary Codes

There is a certain relationship between the missile guidance one-stage code problem presented in Section II of this paper and the error-correcting binary codes that have been extensively studied in information theory. Suppose that one wishes to transmit any one of a large set of messages over a binary channel, and suppose that there is a probability p that a one will be changed to a zero or a zero to a one in the course of transmission. One can encode the messages in such a way that every message differs from every other message in at least four places: for example, 100000, 011100, 111011 and 000111. (If four messages are to be sent, these are the shortest messages possible.) Note that single errors can be corrected immediately (111111 must be 111011), and double errors can be detected but not corrected (110100 can be either 100000 or 011100).

In general, one attempts to encode N messages in lengths as short as possible so that n simultaneous errors in the encoding can be corrected (that is, the original message can be identified). Viewed geometrically, each encoded message has a cluster of closely related correctable messages associated with it (for example, all messages differing from the correct one in only one unit). These clusters are packed as tightly as possible into a binary k-dimensional space (k is the encoded message length) having a total of 2^k points. When the encoded messages are transmitted, synchronization of some sort must be provided between the transmitter and the receiver in order that the receiver may know when each binary message starts. The message length is kept as short as possible in order to maximize the information rate in the channel.*

The missile guidance one-stage codes of this paper can very easily

* For a detailed discussion of these error-correcting binary codes, see Slepian.[1]

be represented as binary messages: each binary digit corresponds to τ time units, and the pulses of the message are represented by ones. For example, the two three-pulse commands (1, 6) and (2, 3) become (11000001) and (101001). To make the representation more precise, the shorter messages can be extended with zeros so that all messages are of equal length.

There are two differences between the missile guidance single-stage codes and the error-correcting binary codes; both are discussed in the paragraphs below. The first difference is concerned with synchronization, and the second with the shapes of the clusters of correctable messages.

A.1 *Synchronization*

The error-correcting binary decoder accepts messages in non-overlapping sets of k digits each (where k is the message length). This implies that some sort of synchronization between the sender and receiver has been established; otherwise the receiver does not know when to start decoding. This synchronization is not easy to provide in missile communications, because of the rapidly changing position of the missile relative to the ground transmitter. Therefore, the missile receiver is arranged to start decoding any time a pulse (that is, a "one") is received.

Some work has been done on self-synchronized codes. For example, Golomb, Gordon and Welch[2] have derived upper bounds for the number of k-digit messages that can be constructed using an n-digit alphabet. They require that the set of messages have the property that the final $(k - i)$ digits of any message in the set followed by the first i digits of any message in the set does not form a message in the set. A set of messages with this property automatically provides synchronization; even when the decoder looks at all the intermediate "messages", it recognizes none of them. However, the work described above has not yet been applied to error-correcting codes. It is possible, in fact, that self-synchronized error-correcting binary codes may be of limited interest to information theorists because of the reduction in the number of messages per second that can be sent over a channel. However, channel capacity and rate of information flow are not objectives of the missile guidance codes.

A.2 *Clusters of Correctable Errors*

The second difference between the missile codes and the binary codes is the shape of the cluster of correctable errors associated with a given message. The binary code cluster usually consists of those k-digit mes-

sages that differ in only a few digits (say, one or two) from the true message. The missile code cluster, on the other hand, is neither "close" to the true message nor easy to describe geometrically. It consists of all k-digit messages that contain ones in n specified places: for example, if 100101 is the message, then 111101, 101101 and 111111 will be decoded as the message. Obviously, the correctable clusters of different messages overlap each other. To separate the clusters as much as possible, the messages are selected so that no member of one cluster with $(2n - 2)$ or fewer ones is a member of any other cluster. [This is simply a restatement of the requirement that $(n - 2)$ false pulses added to any command cannot possibly form a false command.]

In order to visualize these correctable clusters, it is convenient to transform from the k-dimensional discrete binary space of messages to the $(n - 1)$-dimensional continuous space of times between any set of n pulses (true or otherwise). Let us examine the shape of this cluster a little more precisely for n equal to low values. For $n = 2$, the cluster is a line segment 2τ units long* centered on the correct spacing t_1. For $n = 3$, the set of three-pulse messages that will be decoded as a command (t_1, t_2) is contained in the polygon of Fig. 6. This figure corresponds to a delay line that delays the first pulse for a time $t_1 + t_2$, the second pulse for a time t_2, and the third pulse not at all. A set of pulses with original spacing (t_1, t_2) will be brought into exact coincidence by this delay line. The upper right corner of the square has been sawed off because, for original spacings (t_1', t_2') in this region, $t_1' + t_2' > t_1 + t_2 + \tau$, and therefore the first and third pulses will be separated by more than τ after passing through the delay line.

For $n = 4$, all decodable messages are contained in the three-dimen-

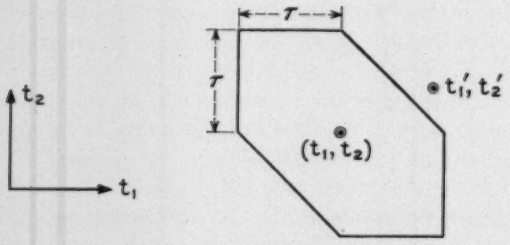


Fig. 6 — Polygon containing set of three-pulse messages that will be decoded as a command (t_1, t_2).

* The quantity τ is the discrete command tolerance defined in Section I. In order for a command to be recognized, the time between the earliest and latest of the n pulses arriving at the AND gate must be less than τ.

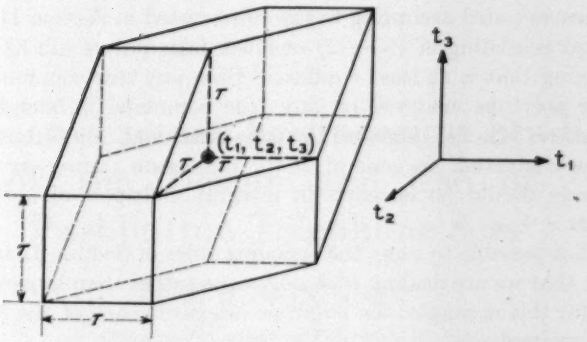


Fig. 7 — Polytope containing all decodable messages for $n = 4$.

sional polytope of Fig. 7. This figure corresponds to a delay line that delays the first pulse $t_1 + t_2 + t_3$, the second pulse $t_2 + t_3$, the third pulse t_3, and the fourth pulse not at all.

For $n \leqq 4$, it can be shown that these polytopes can be packed in such a way as to fill the space completely; it is probably not difficult to show that this is true for arbitrary n. The ratio of the volume of the $(n - 1)$-dimensional polytope to the $(n - 1)$-dimensional hypercube of side 2τ (which encloses it) is

$$\int_0^1 n(n-1)x^{n-2}(1-x)\,dx = \frac{n+1}{2^n}.$$

This integral is equal to the probability that n points drawn at random from a uniform distribution on $(0, 1)$ will all be located within one-half of each other.

To what use can these polytopes be put? In the binary code problem, one places a message at the center of each polytope and packs them into the space as tightly as possible. However, in the missile guidance problem the messages cannot be packed so closely. Consider an n-pulse command plus $(n - 2)$ false pulses. There are $(2n - 2)!/n!(n - 2)!$ possible ways of choosing an n-pulse group, and none of these groups (except for the one consisting of n true pulses) is allowed to lie within any polytope centered on a true command. The polytopes must be very sparsely scattered through $(n - 1)$-space.

Let us assume that the true commands are transmitted with timing errors that are very small with respect to the discrete command tolerance τ. (This is ordinarily true in practice; if it were not, some of the commands might not be received by the missile.) Then, if the true com-

mands are encoded according to the rules stated in Section II, any false command consisting of $(n - 2)$ or fewer false pulses will have at least one spacing that is at least τ different from any true command spacing. But the polytope centered on any true command is bounded by the hypercube of side 2τ; therefore the false command cannot be decoded as any true command. In general, if the maximum timing error is ϵ, the commands should be encoded in integral multiples of a basic time-quantum $\tau + \epsilon$.

It is not possible to relax the encoding rules of Section II to allow for the fact that we are dealing with polytopes rather than hypercubes. The reason for this is simple: the polytope intersects each of the $2^{2(n-1)}$ faces of the hypercube.

REFERENCES

1. Slepian, D., A Class of Binary Signaling Alphabets, B.S.T.J., **35**, 1956, p. 203.
2. Golomb, S. W., Gordon, B. and Welch, L. P., Comma-Free Codes, Can. J. Math., **10**, 1958, p. 202.

Human Factors Engineering Studies of the Design and Use of Pushbutton Telephone Sets

By R. L. DEININGER

(Manuscript received February 16, 1960)

From the user's point of view, what are the desirable characteristics of pushbuttons for use in 500-type telephone sets? The studies reported bear on this question and also on questions of how people process information when keying telephone numbers. Four categories of design features were studied: key arrangement, force-displacement characteristics, button-top design and central office factors. The results indicate that considerable latitude exists for key set design in terms of user performance; however, the preference judgments are more selective. The studies also showed that the manner in which the person acquired and keyed the telephone number influenced performance appreciably.

Technological progress in recent years has brought pushbutton signaling from the telephone set within sight of economic feasibility. What, from the user's viewpoint, are the desirable operating characteristics of the key set which should guide development and manufacture? And how do people process information when they key a telephone number?

I. HUMAN FACTORS PROBLEM

Specifically, we would like to know how pushbutton design influences user speed, accuracy and preference in keying telephone numbers. What design specifications will maximize these three quantities, and how critical is it to achieve these maxima? Also, what other factors influence user information processing in keying telephone numbers? For example: How does performance improve with practice? Does it matter how the number to be keyed is displayed? And are there systematic procedures that users follow in keying numbers?

The design features are discussed in Section III and fall into four groups: (a) key arrangement, (b) force-displacement characteristics, (c)

button top design and (d) central office considerations. Observations concerning other factors in keying behavior are presented in Section IV.

II. EXPERIMENTAL APPROACH

The number of possible key arrangements, force-displacement characteristics and button tops is very large — too large to be tested. A selection of characteristics was, therefore, made on the basis of prior knowledge, user expectation and broad engineering requirements, so that we could examine only the region around an expected maximum.

In general, each series of test sessions extended over three to five days and compared variations of one or two characteristics, with all other characteristics being kept constant. At the end of the tests the preferred values of the individual characteristics were incorporated into a single key set. Evaluation of this set provided a check on the interactions of these individual choices and on how well they fitted together.

It was recognized that people's keying experience with each set during the three to five sessions in a series would be limited compared to the years of practice they could get if pushbutton telephones became a reality. However, methodological studies showed that differences on a relative basis between key set designs appeared after a comparatively small amount of experience.

A group of adjustable pushbutton telephone sets was used in the later studies in the series. To build a telephone set for every change in a characteristic would have been prohibitive in cost. Such changes were simulated by use of specially designed universal pushbutton switches (Fig. 1). Each adjustable telephone set contained ten universal switches mounted in an arrangement determined by the face plate employed.

Typically, a sample of 10 to 15 employees at Bell Telephone Laboratories, Murray Hill, N. J., was drawn randomly for each study. These people came to a laboratory test room and used two to five pushbutton telephone sets that differed in several characteristics, keying 10 to 15 standard telephone numbers on a set each day. A different set was used for each daily session until all the sets in the study had been tried, the order of use being dictated by the design of the experiment. The number of individuals used in the later studies was selected to detect small differences among the average keying times, by considering the power of the analysis of variance tests involved (Ref. 1, p. 379 and 425). The error terms for these tests were reduced by removing the effects of practice and of differences among individuals in the analysis of variance.

On the first day the subjects were told the procedure that would be followed and were asked to key accurately and quickly, as if they were

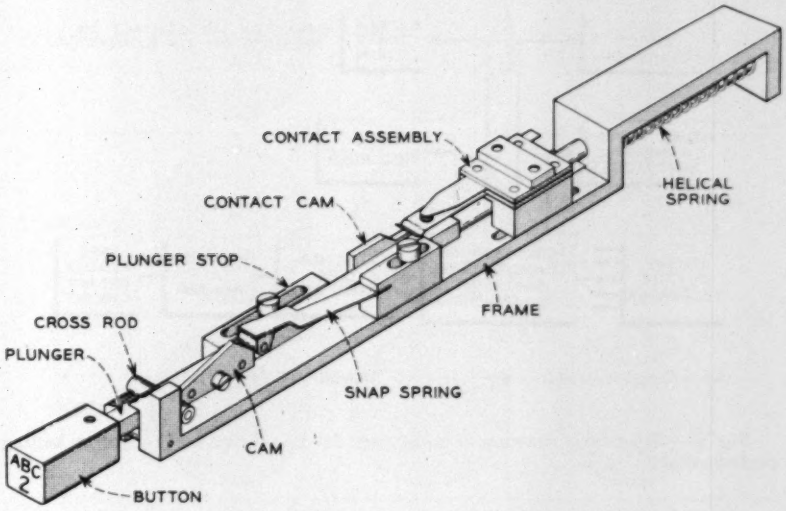


Fig. 1 — The universal pushbutton switch. Changes in force-displacement characteristics were obtained by hand adjustments of the components.

at home or in their office. After trying all conditions in the study, each person was asked for his preferences and for any suggestions he might care to offer. The use of employees in a laboratory-type study seemed justified for two reasons: All comparisons were made within the same study and presumably under constant conditions, and the field trials that followed the laboratory studies would serve to check the findings.

When a person keyed a number, his performance was measured on the equipment shown diagrammatically in Fig. 2. Speed was measured in terms of the keying time, that is, the time interval from the electrical contact of the first key depression until the end of contact of the seventh. Keying accuracy was determined by automatically comparing the number keyed with the number to be keyed. Provisions were made for obtaining more detailed time measures, such as interdigital times and contact durations.

III. THE DESIGN VARIABLES

3.1 *Arrangement of Keys*

The arrangement of the keys was specifically investigated in one study and then checked incidentally in several others. Although other methods

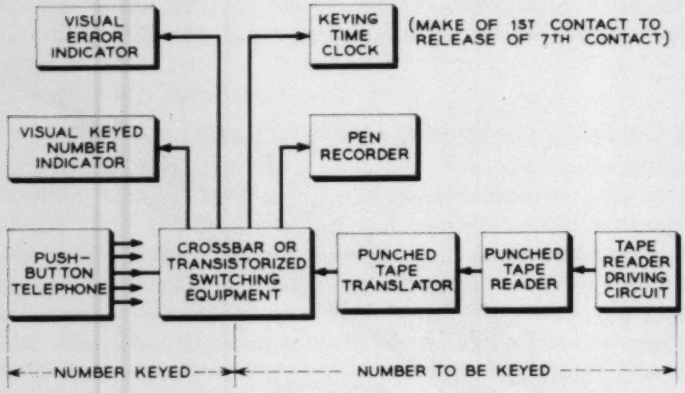


Fig. 2 — Schematic drawing of equipment for measuring and recording keying performance.

were used to make the initial selection of arrangements for the study, performance tests were used to make the final choice because the other methods seemed inadequate. For example, a preliminary study by Lee and Snodgrass[2] showed that there was no significant relation between initial questionnaire preference and subsequent keying performance, and that 10 of the 20 subjects changed their preferences after using the pushbutton telephone sets.

Sixteen different arrangements were selected for the first part of this study. These arrangements were compared by separating them into groups of three and having a different sample of six employees try all three arrangements during each of five sessions. The arrangements and how they were grouped for the study are shown in Fig. 3.

In all the comparisons, only small differences were found in the keying times and errors, and the most preferred arrangements tended to be the best in terms of performance. Significant differences in keying times, errors or preferences were found in four of the six comparisons. Notice that the arrangement frequently found in ten-key adding machines (arrangement I-A, Fig. 3) was not the best of the first three arrangements compared. On the other hand, the same geometric configuration with a different numbering scheme (arrangement IV-A) was superior in keying performance when compared in Group IV. However, the performance differences between the two were small: arrangement I-A had an average keying time of 5.08 seconds, and arrangement IV-A had an average of 4.92 seconds.

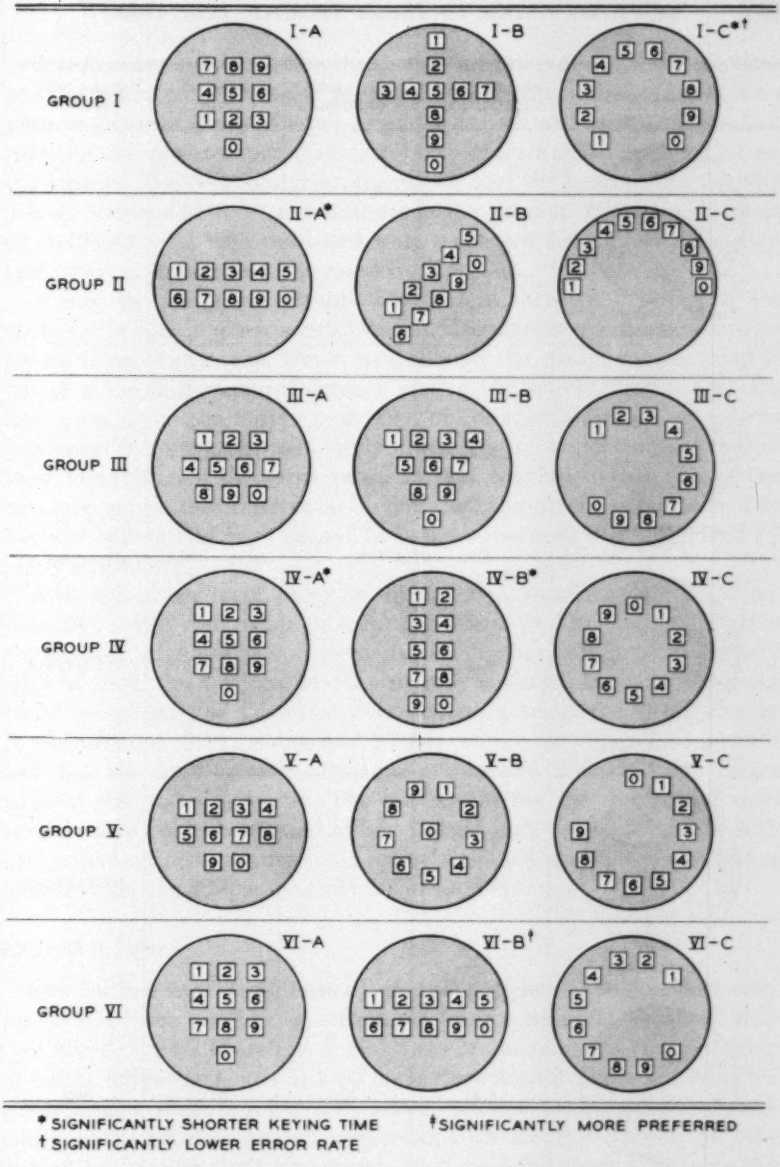


Fig. 3 — The 16 arrangements used in the first study, grouped as they were compared. Two of the arrangements used earlier were compared again in the last group. Although not shown in the figure, the letter groups usually associated with the numbers on a telephone dial were also on the button tops. The tops were $\frac{3}{4}$ inch square with $\frac{5}{64}$-inch-high black letters and numbers on a white background. The circles shown were $4\frac{1}{4}$ inches in diameter.

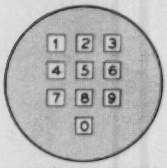
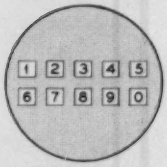
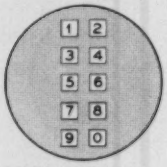
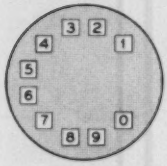
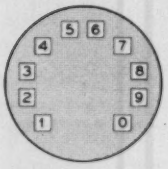
ARRANGEMENT	KEYING TIME (SECONDS)	PER CENT ERRORS	RANKING FOR	RANKING AGAINST
 THREE-BY-THREE PLUS ONE	6.01	2.5	3RD	2ND
 TWO HORIZONTAL ROWS	6.17	2.3	1ST (MOST)	4TH
 TWO VERTICAL COLUMNS	6.12	1.3	5TH (LEAST)	1ST (MOST)
 TELEPHONE	5.90	2.0	2ND	5TH (LEAST)
 SPEEDOMETER	5.97	3.0	4TH	3RD

Fig. 4 — The five arrangements compared in the second study. The specifications listed in the caption of Fig. 3 apply also to this figure.

The four arrangements found superior in their individual comparisons and the arrangement similar to the standard rotary dial (Fig. 4) were used in the next study. In this way, the fastest and most preferred arrangements were compared directly with the standard rotary dial arrangement. The rectangular arrangements had the buttons spaced with $\frac{3}{4}$ inch between centers. The buttons in the circular arrangements were at the ends of $1\frac{1}{2}$-inch radii and were separated by 30°, with the first and tenth buttons being separated by 90°.

A new sample of 15 employees served in three replications of the same Latin square experimental design.[3] No significant differences were found among the keying times even though the study was designed to detect a one-half-second difference among the sets in nine out of ten instances, given the 95 per cent level of confidence and an error term estimated from the previous study. Similarly, no significant differences were found among the error rates. It was concluded that any of the five arrangements was acceptable, but that the arrangement with two vertical columns of keys should be avoided because it was disliked by many subjects.

Although either rectangular or circular arrangements were found acceptable, two of the rectangular arrangements in Fig. 4 offered certain engineering advantages and were studied further. A subsequent study showed that the buttons in the three-by-three-plus-one arrangement could be spaced $\frac{5}{8}$ or $\frac{3}{4}$ inch between centers without significant change in performance (5.56 seconds and 1.7 per cent errors versus 5.54 seconds and 2.5 per cent errors), although preferences indicated the larger spacing was more desirable. The buttons in the two horizontal rows arrangement could be spaced either $\frac{3}{4}$ or $\frac{27}{32}$ inch between centers with little effect on performance (5.33 seconds and 5.8 per cent errors versus 5.63 seconds and 2.5 per cent errors), or on preference.

3.2 *Button Tops*

The button tops used initially in the adjustable pushbutton telephone sets were marginally acceptable from a legibility point of view (see Ref. 4, p. 24). Therefore, a study was conducted to assess the effects of larger letters and button tops on performance and preference. Five combinations of button-top size, letter size and location number size, and pushbutton arrangement were compared (see Fig. 5). Gorton condensed letters and similar numerals were used in the sets, as were the stroke widths recommended by Baker and Grether[4] for the various conditions.

Fifteen randomly selected employees served as subjects in three 5 × 5

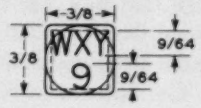
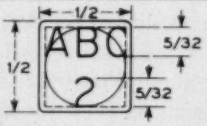
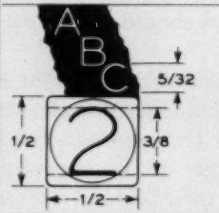
BUTTON ARRANGEMENT	SET AND SPACING	BUTTON AND LETTERING SIZE	REMARKS
(3X3+1)	SET NO. 1: 3/4 IN. BETWEEN BUTTON CENTERS		KEYING TIME: 6.35 SEC KEYING ERRORS: 7.1% 2 VOTES FOR 9 VOTES AGAINST
	SET NO. 2: 3/4 IN. BETWEEN BUTTON CENTERS		KEYING TIME: 5.83 SEC KEYING ERRORS: 1.3% 1 VOTE FOR 1 VOTE AGAINST
(5-5-H)	SET NO. 3: 3/4 IN. BETWEEN BUTTON CENTERS		KEYING TIME: 5.75 SEC KEYING ERRORS: 2.0% 4 VOTES FOR 1 VOTE AGAINST
	SET NO. 4: 3/4 IN. BETWEEN BUTTON CENTERS AND LETTERS ON PLATE		KEYING TIME: 5.77 SEC KEYING ERRORS: 3.3% 6 VOTES FOR 4 VOTES AGAINST
	SET NO. 5: 27/32 IN. BETWEEN BUTTON CENTERS		KEYING TIME: 6.07 SEC KEYING ERRORS: 5.3% 2 VOTES FOR 0 VOTES AGAINST

Fig. 5 — The five conditions compared in the lettering study. Both the "votes for" and "votes against" total to 15, the number of subjects in the study.

Latin squares, this number again being based on the power of the test considerations.

Apparently, there is an optimal size for the button tops. Significant differences were found among the keying times and errors, with the smallest tops and lettering being the poorest. Keying with the middle-size tops was superior to that with the smallest tops, regardless of the arrangement of the keys or the location of the lettering. Keying with the large rectangular tops fell between that with the middle-size and that with the smallest-size tops, but did not differ significantly from either. As is apparent in Fig. 5, the differences among the keying times

were relatively small, even though they were significant: The largest difference was 0.6 out of 6.0 seconds. The error rates differed significantly from one another; however, a large part of the χ^2 was due to the extremes of 1.3 and 7.1 per cent.

The preferences followed the same pattern as the performances. Subjects stated their dislike of the smallest top and lettering and their preference for the middle size. Placing the letters on the plate rather than on the button top was controversial: Six individuals liked the letters off the top, but four others disliked the idea.

The top of the pushbutton seemed to be important because it served as a display for the associated number and letters and as a target for the key-pressing response. Increasing the size of the top and lettering improved the display. Because the over-all size of the keyset was limited, the larger button tops ultimately required a reduction in the separation between adjacent tops, thereby impairing the qualities of the target.

3.3 *Force-Displacement Characteristics*

While the arrangements were being studied, three aspects of the force-displacement curves were under investigation. Prior to that time, preliminary studies showed that gross variations in the feel of the button had little effect on performance. However, these studies showed force in the neighborhood of 100 to 200 grams and displacements of about one-eighth inch were preferred to larger values.

3.3.1 *Force*

In the first study, the force required to depress the button was varied while the maximum displacement was held constant at one-sixteenth inch. Two conditions were used: medium-touch buttons and light-touch buttons (see the dashed curves in Fig. 6). The forces referred to depended on the helical springs used in the button mechanisms, and Fig. 6 shows static force-displacement curves due mainly to the springs.

A sample of 24 employees used each condition for two consecutive days, half trying the medium-touch buttons first. Only small (0.04 second and 0.8 per cent errors) and insignificant performance differences were found. The preference judgments were somewhat obscured by the fact that the subjects were not told the difference between the two sets until the end of the study and did not have the sets in front of them while they made their judgments. Fifteen persons failed to notice a difference, and meaningful preference information could not be obtained from them. However, of the nine who noticed a difference and had a

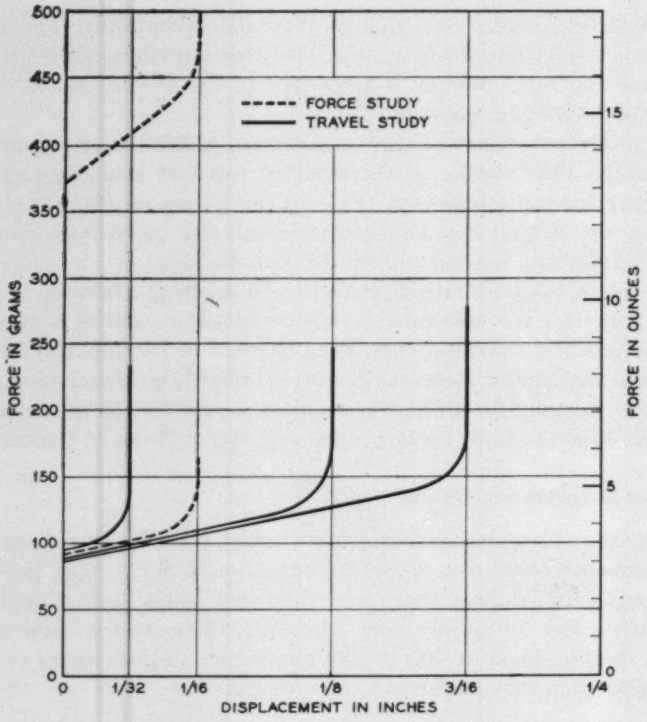


Fig. 6 — The idealized force-displacement curves of the two conditions compared in the study of forces, and the three conditions compared in the study of button travel. These curves are based on essentially static measurements of the universal pushbutton switches.

basis for their preference judgment, eight preferred the light-touch button and one had no preference.

3.3.2 *Travel*

Three conditions of maximum button displacement were compared in the next study. The force was held approximately constant, near 100 grams, and maximum displacements of $\frac{1}{32}$, $\frac{1}{8}$ and $\frac{3}{16}$ inch were compared (see the solid curves in Fig. 6). The point at which the electrical contact was made varied with maximum displacement, so that the contact was closed during 50 per cent of the travel. A new sample of 27 employees tried one condition a day in a cross-over experimental design[3] until all three had been used.

The average keying times ranged from 5.67 to 5.86 seconds and the

error rates from 3.9 to 4.1 per cent. No significant differences were found. When asked which condition they most preferred, 11 subjects said the $\frac{1}{32}$-inch displacement, 9 the $\frac{1}{8}$-inch and 6 the $\frac{3}{16}$-inch. Twelve voluntarily stated they did not like a particular condition: eight disliked the $\frac{1}{32}$-inch, one the $\frac{1}{8}$-inch and three the $\frac{3}{16}$-inch. Taking into account both likes and dislikes, the smallest displacement appears controversial, the largest unpopular and the middle the most desirable.

3.3.3 *Feedback*

The next question studied was as follows: What would happen if additional auditory or kinesthetic-tactile feedback were added to a button incorporation desirable values of force and travel? Letting the customer hear the voice-frequency switching signals might provide feedback concerning the adequacy of the button pressings. Moreover, the addition of a slight snap action and a more distinctive bottoming to the pushbutton might improve performance.

For this study, the buttons in one adjustable set had the force and displacement that were found desirable in the two previous studies. The buttons in a second set had the same travel, but had a slight snap action and a bottoming action that terminated movement abruptly and with an audible click. Each of these sets was used with and without the voice-frequency code signals, so that a total of four conditions were studied. Three 4 × 4 Latin squares were used with 12 employees, who tried one condition a day.

For both performance and preference, the differences among the four conditions were small and insignificant. This study is interpreted as indicating that neither form of additional feedback was necessary in a button mechanism that had desirable force-displacement characteristics.

3.3.4 *Composites of Characteristics*

Because the studies reported thus far investigated one or two of the characteristics at a time, there remained the question whether performance and preference for composites of the characteristics could be predicted from the individual studies. This point was checked in the following study that compared three pushbutton telephone sets, each embodying a different composite of characteristics.

The preferred characteristics isolated by the preceding studies were combined in one of the adjustable pushbutton sets. The other two sets in this study had been designed for various engineering and test purposes

before the preferred values were isolated. When the characteristics of these two sets were assessed in terms of the individual human factors studies, it was found that they deviated from the most preferred values but were still largely within the range of desirable and acceptable values. The largest deviations were found in the force-displacement characteristics, but even these were not extreme. Thus, it was predicted that (a) small and insignificant performance differences would be found, but (b) large and significant preference differences would be found, particularly where the force-displacement characteristics of the button mechanism were concerned.

Forty-five employees, none of whom had served in an earlier study, used one set each day to accord with 15 3 × 3 Latin squares. This sample size was selected to provide a more powerful test of the error rates than was given by the earlier studies and to permit more detailed analysis of the preference judgments.

As predicted, the differences among the average keying times were small and insignificant. An average of 5.8 seconds was required to key a standard telephone number on the most preferred combination versus 5.9 and 6.0 seconds on the other two. However, significant differences were found among the error rates because the electrical contacts on one set were out of adjustment. At least one error was made in keying 2.3 per cent of the telephone numbers on the most preferred composite, in keying 2.0 per cent of the numbers on the second set and in keying 10 per cent of the numbers on the third set. The additional 8 per cent errors were all of one type; when the fault was corrected in later models, more normal error rates were obtained with that set.

The large differences expected among the preference judgments for the three composites were found. In an interview after the last day of testing, the subjects were asked which set they preferred the most and which they preferred the least. Although there were no significant differences among the first place votes, the set with the most preferred characteristics was rated as "least liked" by significantly fewer people than either of the two other sets. The subjects were then asked which button feel they liked the most and which they liked the least. Again, the set with the most preferred characteristics came out ahead: not a single person disliked the feel of its mechanisms.

In a sense, this study served to validate the procedure adopted in this series, since it showed that the desirable characteristics isolated in the individual studies could be combined into a superior set. Very likely, this was because much the same procedure was used throughout the entire series. Also, the fact that the characteristics of the other two sets

were close to the characteristics investigated in the individual studies facilitated agreement between the predictions and the results.

3.4 *Central Office Factors*

How long the customer holds the pushbutton in contact and how long he pauses between consecutive key pressings are important in the design of central office switching equipment. Of primary concern are short contact durations and brief interdigital times. Some data regarding these two intervals were gathered in two studies.

Fifteen employees keyed a total of 1500 seven-character all-numeral numbers on each of two pushbutton telephone sets in the first study. These sets were prototypes of operating equipment and differed in three characteristics: the arrangement of the keys, the force required to depress the keys and the point at which the electrical contact was made as the button was being depressed.

Keying was about as fast on one set as on the other; however, the contact durations were shorter and the interdigital times longer on one set than on the other (see Fig. 7). This differential proportioning of the time seems to reflect the relative location of the electrical contact in the travel of the button as indicated in the figure. The results of this study agree with the results of a preliminary study in which five prototype key sets were studied. Thus, it may be possible to lengthen the contact duration by increasing the percentage of button travel during which contact is made.

Very brief contact durations did occur; for example, durations equal to or less than 0.042 second occurred in 1 per cent of the key pressings. The interdigital times were longer, and 1 per cent of them fell at or below 0.090 second. The occurrence of short time values depended on the person doing the keying, particularly for the interdigital times, where most of the short values were due to one individual. Another source of

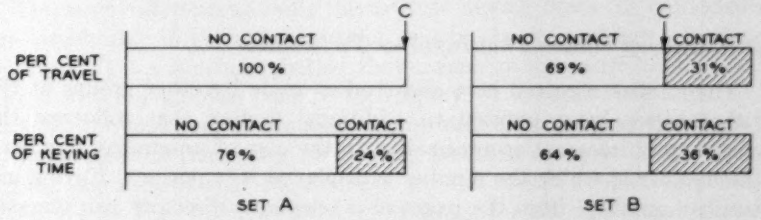


Fig. 7 — The effect of the location of the electrical contact on the percentage of keying time during which contact was made.

short interdigital intervals was a repeated character in the number to be keyed. The interdigital time for repeated characters averaged 0.100 second less than the times for nonrepeated characters.

A second set of detailed measurements, taken during part of the study of composites, showed more clearly that brief contact durations and interdigital times were associated with fast keyers. The last ten keyings made by 30 of the 45 individuals using one of the prototype sets were recorded, equipment permitting. The correlations among the average contact duration, the average interdigital time and the average keying time for each subject were computed after the averages were transformed logarithmically to reduce the skewness of the distributions. The average keying time correlated 0.47 with the average contact duration and 0.76 with the average interdigital interval. No relation was found between the contact duration and the interdigital time. The correlation coefficients indicate that fast keyers tend to have shorter contact durations, and definitely make shorter interdigital pauses. This latter point will be discussed in greater detail in a moment, for it seems related to the procedure that each subject adopted in keying numbers.

IV. OBSERVATIONS ON KEYING BEHAVIOR

When a person keys a telephone number he is processing information in a very literal sense. The standard telephone number contains 22.3 bits of information; on the basis of the key pressings, the central office switching equipment selects one line and completes the call. The physical characteristics of the pushbutton key set are one factor that influences the efficiency of the information processing. As was reported in Section III, the speed and accuracy of performance were affected by the arrangement of the keys and the size of the button tops. Moreover, if larger forces and longer displacements were studied, or if more powerful experiments were used, significant effects very likely would be found for the force-displacement characteristics as well.

4.1 *Additional Factors That Influence Performance*

The studies reported here and studies made by other groups at the Laboratories have pointed to additional factors that influence the efficiency of information processing. In the case of unfamiliar numbers, the manner in which the number is displayed is important. Keying unfamiliar numbers from the pages of a telephone directory can increase the keying times by 75 per cent and the keying errors by 100 per cent in comparison with keying from a 3- × 5-inch file card on which only

one number is typed. The amount of information to be processed is also a factor; for example, about 4 seconds more is required to key a seven-character telephone number than a four-character one. Familiar numbers and numbers with repeated characters or simple sequences of digits are keyed quickly and accurately.

Experience with pushbutton keying facilitated the information processing. Keying became faster as people used the key sets day after day. However, performance seemed to improve at about the same rate whether they used the same set each day, the same two or three sets each day, or a different set each day. Some of this improvement was due to increased familiarity with the test room and procedure, but similar improvements with practice appear in field trials, indicating that experience in using pushbuttons is important. In the case of keying accuracy, the low error rates make detection of any learning trends difficult.

Perhaps the most important factor in the information processing is the individual himself. Some people keyed seven-character telephone numbers in less than two seconds, requiring about 2.7 seconds on the average, and others required as high as 12.4 seconds on the average to key the same numbers. These differences were consistent from day to day and from telephone to telephone, as can be gathered from the following rank-order correlation coefficients: There was a 0.8 correlation between pushbutton keying times on day 1 and pushbutton keying times on days 12 through 14, a 0.7 correlation between pretest rotary dialing times and pushbutton keying times on day 1 and 0.6 correlation between pretest dialing times and pushbutton keying times on days 12 through 14. The keying times on two different pushbutton sets were more highly correlated than rotary dialing times and pushbutton keying times.

4.2 *The Importance of the Keying Process*

The large differences among the average keying times for individuals are largely due to the ways in which people acquire and key telephone numbers. It is a question whether they memorize the entire number and key it without referring back to the display, or whether they memorize and key the first part of the number and then refer back to the display to memorize the remainder for keying. Detailed analysis of the interdigital times shows that the intervals for the fastest subject vary only slightly (see Fig. 8). The interdigital times for average-speed keyers show a distinct increase between the third and fourth characters. This accounts for the fact that the average keying time is more highly cor-

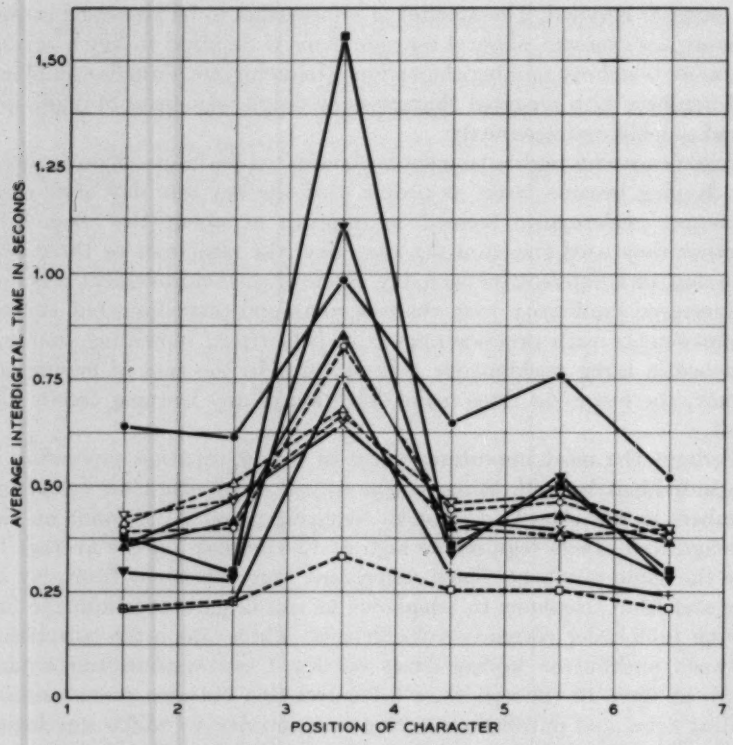


Fig. 8 — Average interdigital intervals for individual subjects keying seven-character numbers. The interdigital intervals were measured from the breaking of one electrical contact to the making of the next contact. (These results were obtained for all-numeral dialing; data for letter-and-numeral dialing are similar but less definitive than those shown in the figure.)

related with the interdigital time than with the contact duration. These findings corroborate the statements made by people about how they key telephone numbers. The fastest say that they memorize and key the entire number without referring back to the display, whereas the average to slow keyers say that they refer back in the middle of keying the number.

4.3 *Why Refer Back?*

Neither method of keying seems more accurate than the other. One might predict that keying without referral would be less accurate and therefore that fast keyers would make more errors than slow keyers. However, no relation between keying time and errors could be detected

among the studies reported, regardless of whether they were considered singularly or as a group for analysis. The absence of a correlation between time and errors in the present studies must be interpreted cautiously, since these studies were not designed to detect such relations.

On the other hand, referral does serve an important purpose, as indicated by the following study. A sample of six employees was asked to key telephone numbers with referral during two sessions and without referral during two other sessions. When keying without referral, the individuals were asked to memorize the entire number, turn over the display card, and key the number from memory. A total of 156 standard telephone numbers were keyed under each set of instructions. Some subjects found it very difficult to key telephone numbers without referral. These people said they usually dialed unfamiliar telephone numbers by breaking them into two or more parts. The difficulty they experienced was reflected in their error rates. There was a drop in keying accuracy as more and more of the characters in the number were keyed, and this drop was due primarily to those who said they usually dialed with referral (see Fig. 9). Apparently, if a person habitually keys with referral,

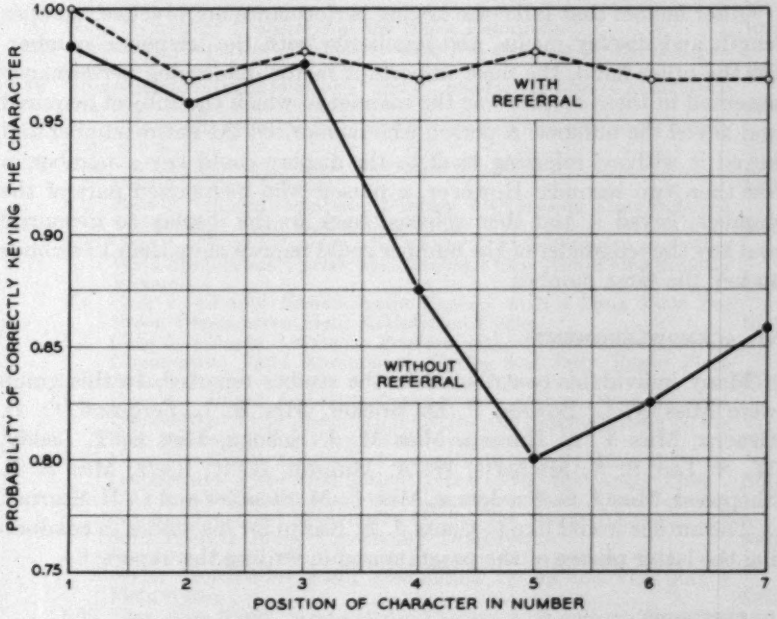


Fig. 9 — Accuracy of keying as a function of position of character for two methods of keying. The number display was always present during keying with referral, but it was removed before keying started for keying without referral.

he does so to increase the accuracy of keying the latter characters in the telephone number, even though his over-all keying accuracy is not appreciably greater than that of a person who usually keys without referral.

V. CONCLUSIONS

Regarding the design of pushbutton telephone key sets that are fast, accurate and convenient to use, the following statements can be made:

i. The operating characteristics of the key sets significantly influence both keying performance and user preference.

ii. In terms of keying performance there exists a rather broad region of desirable values for the operating characteristics. Thus, there is latitude for telephone key sets and, so long as the characteristics remain in the region of desirable values, little deterioration in keying performance will be found.

iii. Considerably less latitude exists if preferences are considered, particularly in the case of the force-displacement characteristics. Typically, subjects preferred a smooth and quietly operating button with a light touch and a moderate travel.

Other factors that influence keying performance are practice, number length and display media, and familiarity with the telephone number. On the other hand, the most important factor influencing performance observed in these studies was the manner in which the subject acquired and keyed the number. A person who memorized the entire number and keyed it without referring back to the display could key a number in less than two seconds. However, a person who memorized part of the number, keyed it and then referred back to the display to memorize and key the remainder of the number could require more than 12 seconds to key the same number.

VI. ACKNOWLEDGMENTS

Many individuals contributed to the studies reported. In this group were Miss N. L. Bowles, P. D. Bricker, Mrs. S. L. Ferguson, O. O. Gruenz, Miss V. A. Hansen, Miss M. J. Kellogg, Miss E. T. Leddy, W. A. Lee, S. E. Michaels, W. A. Munson, R. R. Riesz, Mrs. S. B. Sheppard, Miss J. G. Snodgrass, Miss C. M. Steadler and C. H. Sturner.

The author would like to thank J. E. Karlin for his advice in conducting the latter phases of the program and in writing this report.

REFERENCES

1. Hald, A., *Statistical Theory with Engineering Applications*, John Wiley and Sons, New York, 1952.
2. Lee, W. A. and Snodgrass, J. G., On the Relation Between Numbering Preferences and Performance on a Ten-Button Keyboard, Amer. Psychol., **13**, 1958, p. 425.
3. Federer, W. T., *Experimental Design*, Macmillan, New York, 1955.
4. Baker, C. A. and Grether, W. F., Visual Presentation of Information, Wright Air Development Center, TR 54-160, 1954.

Passage Effects in Paramagnetic Resonance Experiments

By M. WEGER

(Manuscript received February 16, 1960)

An attempt is made to classify theoretically paramagnetic resonance signals of inhomogeneously broadened lines occurring under various experimental conditions. The theoretical predictions are checked experimentally. Special emphasis is given to cases in which T_1 is long, and to cases in which the adiabatic condition $\gamma H_1^2 \gg dH/dt$ is violated.

TABLE OF CONTENTS

I. Introduction 1014
II. Some Symmetry Properties of the Bloch Equations 1015
III. Symmetry Properties of Some Systems Not Obeying the Bloch Equations 1020
IV. Approximate Expressions for χ' and χ'', and for the Loss of Magnetization for a Single Spin Packet 1028
V. Approximate Expressions for χ' and χ'' for an Inhomogeneously Broadened Lines 1030
VI. The Various Passage Cases — Detailed Discussion 1031
6.1 Case 1 — Slow Passage 1033
6.2 Case 2 — Rapid Adiabatic Passage with a Long Time Between Consecutive Field Modulation Cycles 1035
6.3 Case 3 — Rapid Passage with a Long Time Between Consecutive Field Modulation Cycles, the Adiabatic Condition Being Slightly Violated 1040
6.4 Case 4 — Rapid Nonadiabatic Passage with a Long Time Between Consecutive Field Modulation Cycles 1047
6.5 Case 5 — Rapid Adiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles and Very Rapid Field Sweep 1048
6.6 Case 6 — Rapid Adiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles, the Magnetization of the Spin Packets Being Partly Destroyed 1052
6.7 Case 7 — Rapid Adiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles — the Stationary Case 1056
6.8 Case 8 — Rapid Adiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles, the Magnetization of the Spin Packets Being Completely Destroyed 1060
6.9 Case 9 — Rapid Nonadiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles and Very Rapid Field Sweep 1063
6.10 Case 10 — Rapid Nonadiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles — the Stationary Case 1064

6.11 Case 11 — Rapid Nonadiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles, the Magnetization of the Spin Packets Being Completely Destroyed 1067
6.12 Conditions to Maximize Signal 1070
VII. The Various Passage Cases — Diagrammatic Representations and Additional Experimental Points 1072
VIII. Conclusion 1078
IX. Acknowledgments 1081
X. Notation 1081
Appendix A. χ' and χ'' for a Single Spin Packet, for Long Relaxation Times 1083
Appendix B. χ' and χ'' for an Almost Sudden Passage of an Inhomogeneously Broadened Line, Employing a System with a Bandwidth Narrow Compared with the Line Width 1091
Appendix C. χ' and χ'' for a Nonadiabatic Passage, Nonuniform Rate of Sweep 1096
Appendix D. Superposition of Adiabatic Fast Passage Lines 1103
Appendix E. Second-Order Effects Occurring When a Bridge Is Tuned to Observe χ' or χ'' Signals 1107
Appendix F. A Particular χ' Trace Observed at High Powers 1111
References 1112

I. INTRODUCTION

Shapes of electron spin resonance (ESR) and nuclear magnetic resonance signals, observed under various conditions, have previously been predicted theoretically and confirmed experimentally for a few cases.[1-5]

The purpose of this work is as follows:

i. To understand and predict the shapes of observed traces, as an aim in itself.

ii. To use the shapes of the traces to gain some understanding about the local fields in the sample, mainly to understand to what extent the fields can be considered static, and to what extent they must be considered as dynamic (namely, responsible for forbidden lines, etc.).

iii. To determine various conditions under which the resulting traces may cause misleading determinations of line shapes, relaxation times, etc.

iv. To distinguish true "physical" effects from spurious "passage effects" (like splitting of lines).

The work was performed on single crystals of phosphorus-doped silicon, containing about 1.7×10^{16} donors per cm, at temperatures of 10°, 4.2° and 1.2°K. The ESR line due to the donors was observed at about 9300 mc, employing a standard superheterodyne microwave spectrometer,[6] under various experimental conditions. (Mainly various rates and periods of the modulating magnetic field, and various microwave power levels.) Various traces were predicted theoretically and observed experimentally (or vice versa). No attempt was made to measure the intensities of signals absolutely.

The shapes of the traces under various experimental conditions were predicted theoretically by Portis.[5] It was originally attempted to con-

firm those predictions, but it was soon realized that some additional experimental factors must be taken into account — such as the loss of magnetization during adiabatic fast passage or nonadiabatic conditions. However, the basic results of Portis have been confirmed experimentally.

This work does not attempt to be complete or rigorous, but rather to point out some of the most salient effects of various experimental conditions upon observed electron spin resonance lines.

II. SOME SYMMETRY PROPERTIES OF SYSTEMS OBEYING THE BLOCH EQUATIONS

Some spin systems obey the Bloch equations,[1] which in the rotating frame are:

$$\frac{dS_x}{dt} + \frac{S_x}{T_2} - \gamma H_1 \delta S_y = 0,$$

$$\frac{dS_y}{dt} + \frac{S_y}{T_2} + \gamma H_1 \delta S_x - \gamma H_1 S_z = 0,$$

$$\frac{dS_z}{dt} + \frac{S_z}{T_1} + \gamma H_1 S_y = \frac{S_0}{T_1},$$

where S_x, S_y, S_z are the components of the magnetization in a coordinate system, rotating about an externally applied magnetic field H, with S_z being parallel to H. The angular frequency of this rotation is ω. In these equations, $\delta = (H - H_0)/H_1$, where H_0 is the magnetic field at which the system rotates at an angular frequency ω, and H_1 is the magnetic field along S_x, in the rotating frame; $\gamma = \omega/H_0$ is the gyromagnetic ratio of the spin system ($\gamma \approx 1.7 \times 10^7$ rad/sec/gauss for electrons); and T_1 and T_2 are constants known as relaxation times. Let

$$\chi' \equiv \frac{S_x}{2H_1}, \qquad \chi'' \equiv \frac{S_y}{2H_1}.$$

In this work, it will be assumed that the spin systems considered obey the Bloch equations (these will be called "homogeneously broadened lines" or "spin packets"), or consist of noninteracting subsystems that obey the Bloch equations (these are called "inhomogeneously broadened lines"), unless the contrary is explicitly stated.

The behavior of systems obeying the Bloch equations under the transformation $\delta \rightarrow -\delta$; $(dH/dt \rightarrow -dH/dt)$ has been discussed by Jacobsohn and Wangness.[3] The considerations are as follows:

$$\text{If } \delta \rightarrow -\delta, \qquad S_x \rightarrow -S_x, \qquad S_y \rightarrow S_y, \qquad S_z \rightarrow S_z,$$

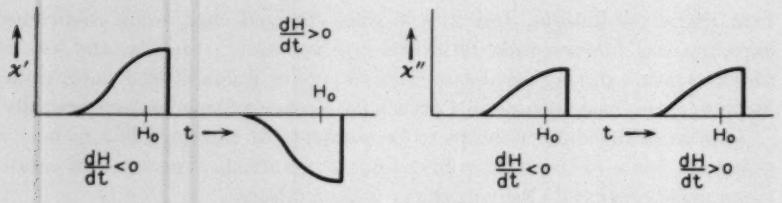


Fig. 1 — Symbolic indication of the parities of χ' and χ'' signals of systems obeying the Bloch equations. Under $dH/dt \to -dH/dt$, we have: $\chi'(H - H_0) \to -\chi'(H_0 - H)$; $\chi''(H - H_0) \to \chi''(H_0 - H)$.

the Bloch equations remain invariant. Therefore, if we start sweeping the magnetic field from an equilibrium state ($S_z = S_0$) not on the line, towards the line, if we reverse the direction of sweep, χ' changes sign, ($\chi' = S_x/2H_1$; $S_x \to -S_x$), while χ'' does *not* change its sign (since $S_y \to S_y$, $\chi'' = S_y/2H_1$ — see Fig. 1).

The change in sign of χ' is well known in the "slow passage"† and "adiabatic rapid passage"† cases [see Figs. 2(a) and 2(b)]. The lack of change of sign of χ'' is also well known in the "slow passage" case [see Fig. 2(c).] Actually, if χ'' did change sign, it would mean that power could be extracted from the system.

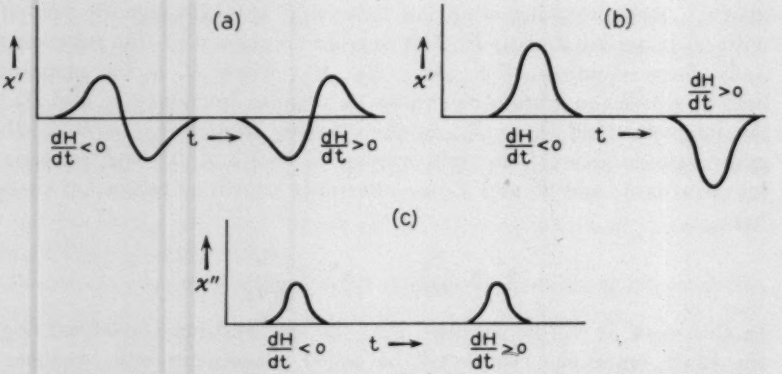


Fig. 2 — The most commonly observed χ' signals of systems obeying the Bloch equations, for (a) slow passage and (b) rapid passage; note the symmetry under $dH/dt \to -dH/dt$. (c) The most commonly observed χ'' signals of a system obeying the Bloch equations, slow passage.

† The "slow passage" and "adiabatic rapid passage" cases are discussed by Bloch.[1] In the slow passage case, $\delta \ll 1/\sqrt{T_1T_2}$, and in the adiabatic rapid passage case, $\delta \gg 1/\sqrt{T_1T_2}$, $\gamma H_1 \gg \delta$.)

If the χ' and χ'' signals are observed by means of a phase sensitive detector employing magnetic field modulation (Bloch[1]), the symmetry property will be just reversed. This occurs because the magnetic field H is given by

$$H = H_0 + \left(\frac{dH}{dt}\right)_0 t + H_m \cos \omega_m t,$$

where H_0, $(dH/dt)_0$, H_m and ω_m are constants, known as dc field, sweep rate, modulating field and modulation angular frequency, respectively.

Under the transformation $\delta \rightarrow -\delta$,

$$H \rightarrow H_0 - \left(\frac{dH}{dt}\right)_0 t - H_m \cos \omega_m t.$$

The phase of the modulating field with respect to the reference signal $\cos \omega_m t$ is inverted. Thus, if we start sweeping from a relaxed state, the sign of the component of χ' at the modulating frequency (relative to the reference voltage), is the *same* for sweeping in both directions, while the sign of χ'' is reversed.†

$$\text{If } \delta \rightarrow -\delta, \qquad S_x \rightarrow S_x, \qquad S_y \rightarrow -S_y, \qquad S_z \rightarrow -S_z,$$

the Bloch equations also remain invariant.

Thus, if S_z is *inverted* during the passage, on returning, χ' will maintain its sign and χ'' will reverse it.

The opposite sign of χ'' (emission) is not absurd, since the level populations have been inverted and we may have a "maser" effect.

Again, with a phase-sensitive detector (PSD), χ' will invert its sign and χ'' will maintain its sign. For some typical traces, see Figs. 3 and 4. (Some of these shapes will be discussed in more detail later.) Table I shows the parity of χ' and χ'' signals, summarizing these results for systems obeying the Bloch equations.

The preceding symmetry properties apply to both homogeneously and inhomogeneously broadened lines. This point deserves further comment. There is no reason why an assembly of noninteracting subsystems must possess any kind of symmetry under a transformation affecting the magnetic field. There need not be any correlation between subsystems resonating at different magnetic fields. However, often a system con-

† Note that, if the second harmonic signal is observed by applying the second harmonic of the field modulation signal as reference to the phase sensitive detector, then inverting the sign of the field modulation voltage will *not* change the sign of the second harmonic reference voltage. Thus, under the above transformation, the second harmonic signal of χ' is reversed, while that of χ'' is not. The same applies to all even harmonics. The odd ones behave like the fundamental.

sists of "almost" identical subsystems, differing only in the distribution of local magnetic fields. The magnetic fields at which such systems resonate, (at a given frequency) are usually distributed symmetrically about some center. In that case, the symmetry properties just discussed are retained. Therefore, if the observed signals do *not* possess a definite parity, this must be due to one of the following four causes:

Systems *obeying* the Bloch equations, or consisting of noninteracting subsystems obeying them, may possess

i. an asymmetrical envelope of an inhomogeneously broadened line;

ii. partial relaxation between the times the line is traversed in opposite directions (neither $S_z \rightarrow S_z$ nor $S_z \rightarrow -S_z$);

iii. bad experimental adjustment of the equipment — for instance, a mixture of χ' and χ'' signals. (Note that, if the signals are not small, the

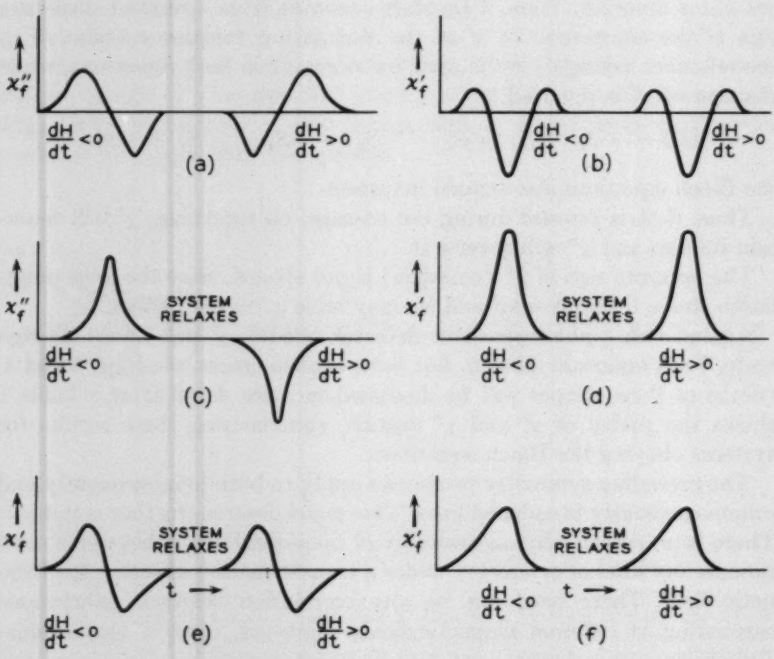


Fig. 3 — Typical traces observed when field modulation and a phase sensitive detector is used: (a) χ''_f, slow passage; (b) χ'_f, slow passage; (c) χ''_f, rapid (non-adiabatic) passage; (d) χ'_f, rapid (adiabatic) passage; (e) χ'_f, rapid passage, very fast weep; (f) χ'_f, a line with spin diffusion (a system *not* obeying the Bloch equations). Note the symmetry under $dH/dt \rightarrow -dH/dt$.

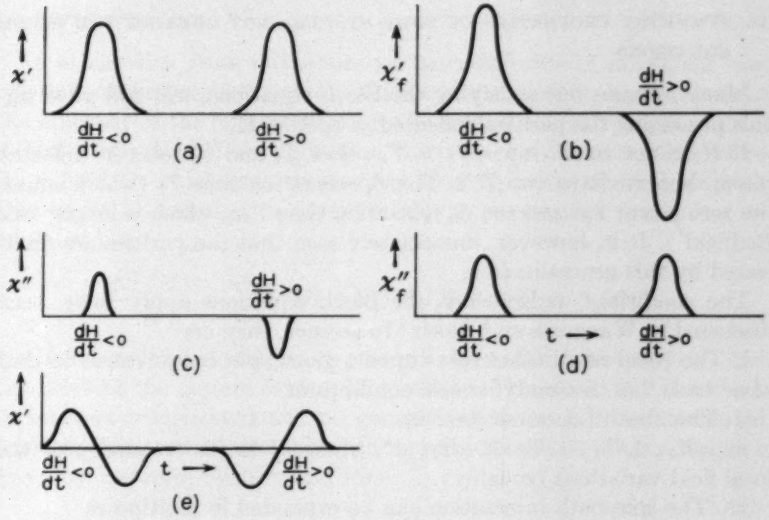


Fig. 4 — Some typical traces observed when the system does *not* relax between consecutive sweeps: (a) χ', *no* phase sensitive detector; (b) χ', *with* P.S.D.; (c) χ'', *no* P.S.D. (rapid, not perfectly adiabatic, passage); (d) χ'', *with* P.S.D.; (e) χ', *with* P.S.D., very rapid sweep. Again, notice the symmetry properties which are *different* from those occurring when the system relaxes between sweeps.

first order bridge theory does not apply; this point will be discussed in Appendix E.)

Systems *not* obeying the Bloch equations may possess

iv. A "genuine" asymmetry.

Some systems of this kind will be discussed in the next section.

TABLE I — PARITY OF χ' AND χ'' SIGNALS UNDER THE TRANSFORMATION $\frac{dH_0}{dt} \rightarrow -\frac{dH_0}{dt}$

		Direct Signal (and Even Harmonics)	Signal as Detected by Phase Sensitive Detector (and Odd Harmonics)
χ'	System relaxes (between sweeps)	−	+
	System does not relax (between sweeps); adiabatic conditions	+	−
χ''	System relaxes (between sweeps)	+	−
	System does not relax (between sweeps); adiabatic conditions	−	+

III. SYMMETRY PROPERTIES OF SOME SYSTEMS NOT OBEYING THE BLOCH EQUATIONS

Many systems not satisfying the Bloch equations, will still yield signals possessing the parities indicated in Section II.

If H_1 is not small, say, $\gamma H_1 > T_2$, then S_x and S_y relax at different rates; thus, we have two T_2's: The S_y relaxation time T_2 (which equals the zero power T_2) and the S_x relaxation time T_{2s}, which is longer (see Redfield[7]). It is, however, immediately seen that the parities are unaffected by this generalization.

The conditions under which the Bloch equations apply have been discussed by Wangness and Bloch.[8] In essence, they are:

i. The effect of all other spins upon a given spin is equivalent to that of a "bath," in thermodynamical equilibrium.

ii. The "bath" does not heat up.

iii. γH_1, $1/T_1$, $1/T_2 \ll \gamma H_0$, ω^*, where ω^* is the frequency of the local field variations (roughly).

iv. The spin-bath interaction can be expanded in multipoles.

v. Either $s = \frac{1}{2}$,
or $s = 1$, $kT \gg \hbar\omega_0$;
or $s > 1$, $eqQ \ll \hbar\omega_0$;
or $s > 1$, $\omega^* > \omega$, isotropic interaction
(eqQ is the quadrupole interaction).

3.1 *Cross Saturation Within a Line*

Often a line is neither homogeneously broadened nor inhomogeneously broadened, but rather something intermediate: the system consists of subsystems which obey the Bloch equations, except for a (not too strong) interaction among them. Such systems have been discussed by Bloembergen *et al.*[9] Often it may be possible to describe the cross relaxation by means of a diffusion equation like

$$\frac{\partial S_z(H)}{\partial t} = \text{constant} \times \frac{\partial^2 S_z(H)}{\partial H^2},$$

which is linear in S_z and invariant under $H \rightarrow -H$, and thus does not perturb the parity.†

Therefore, one would not expect cross relaxation to destroy the symmetry properties of χ' and χ'' discussed in the preceding section.

† Bloembergen's rate equations are nonlinear, but homogeneous in the occupation numbers (for an n-spin flip), and, for an inhomogeneously broadened line with a symmetrical envelope, they are invariant under $\delta \rightarrow -\delta$. Therefore, the above mentioned parity is maintained.

3.2 *Forbidden Lines*

In a line with cross relaxation, we have diffusion of excitation from one part of the line to the other, which is not explicitly dependent upon the intensity of the microwave magnetic field. In some cases, the rate of diffusion will depend strongly upon the intensity of the applied microwave field, in which case we may say that we have "forbidden lines." One might be tempted to describe such a process with the aid of an equation like

$$\frac{\partial S_z(H)}{\partial t} = \sum_{\delta H} a(\delta H, \gamma H_1)\,[S_z(H + \delta H) - S_z(H)].$$

However, as the following examples will show, this equation may often not describe the system even approximately. Therefore, this case seems to be more complicated and no general discussion will be attempted. Instead, three special examples will be considered. (Another example is given by Bardeen, Slichter and Pines.[10])

3.3 *Example 1*

Consider a system consisting of pairs of electrons and (spin $\frac{1}{2}$) nuclei, with dipolar and contact interaction between them.

Let

s_e = spin of electron,
s_n = spin of nucleus,
μ_e = magnetic moment of electron,
μ_n = magnetic moment of nucleus,
H = magnetic field (in the z-direction),
a = contact interaction,
b = dipolar interaction,
r = electron-nucleus distance.

Then, the Hamiltonian of the system is

$$\mathcal{H} = \frac{\mu_e H s_{z_e}}{s_e} + \frac{\mu_n H s_{z_n}}{s_n} + a(\mathbf{s}_e \cdot \mathbf{s}_n) + b\left(\mathbf{s}_e \cdot \mathbf{s}_n - \frac{3(\mathbf{s}_e \cdot \mathbf{r})(\mathbf{s}_n \cdot \mathbf{r})}{r^2}\right).$$

Let the eigenfunctions of the electron spin corresponding to the eigenvalues $s_z = +\frac{1}{2}$, $s_z = -\frac{1}{2}$, be

$$\left|\uparrow\right> \text{ and } \left|\downarrow\right>,$$

respectively, and those of the nucleus be $|\uparrow>, |\downarrow>$. Then, to the

zeroth order (for $\mu_e H \gg \mu_n H$, a, b), the eigenfunctions of the combined system are

$$\left|\uparrow\uparrow\right>, \quad \left|\uparrow\downarrow\right>, \quad \left|\downarrow\uparrow\right>, \quad \left|\downarrow\downarrow\right>.$$

The energy level diagram is shown in Figs. 5 and 6. Under a variable magnetic field H_1 perpendicular to H, the transitions

$$\left|\uparrow\uparrow\right> \leftrightarrow \left|\downarrow\uparrow\right> \text{ ("line 1")}$$

and

$$\left|\uparrow\downarrow\right> \leftrightarrow \left|\downarrow\downarrow\right> \text{ ("line 2")}$$

are "allowed", i.e., occur even for $a = b = 0$; while the transitions

$$\left|\uparrow\uparrow\right> \leftrightarrow \left|\downarrow\downarrow\right> \text{ ("line 3")}$$

and

$$\left|\uparrow\downarrow\right> \leftrightarrow \left|\downarrow\uparrow\right> \text{ ("line 4")}$$

are forbidden, i.e., occur only if $b \neq 0$. Both forbidden lines need not be of the same intensity. Assume temporarily, for the sake of the argument, that line 4 is strictly forbidden, while line 3 is only partly forbidden.

The populations of the levels at thermal equilibrium will be, approximately $1 + \epsilon$ for

$$\left|\downarrow\downarrow\right> \text{ and } \left|\downarrow\uparrow\right>,$$

and $1 - \epsilon$ for

$$\left|\uparrow\uparrow\right>, \quad \left|\uparrow\downarrow\right>,$$

where $\epsilon = \mu_e H/kT$.

If we sweep from a *low* magnetic field, increasing it, line 3 will be swept first, inverting the populations of the corresponding levels (for adiabatic rapid passage). Then, if no relaxation takes place during the sweep, no more lines will be observed, since for lines 1 and 2, the populations of the "upper" and "lower" levels are equal.

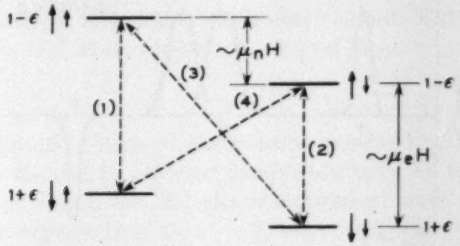


Fig. 5 — Energy-level diagram of an electron-nucleus system in a magnetic field (the nuclear spin is supposed to be ½). Lines (1) and (2) are "allowed," while lines (3), (4) are "forbidden," but not strictly.

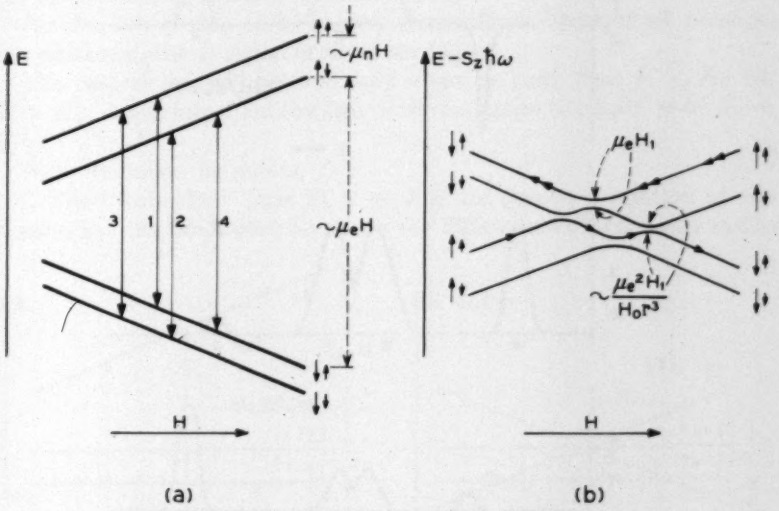


Fig. 6 — The energy levels of an electron and spin half nucleus, as function of the magnetic field for (a) fixed frame and (b) rotating frame. The different behavior of a system in the state $\left|\uparrow\uparrow\right>$ under $dH/dt > 0$ and $dH/dt < 0$ can be seen.

If we sweep from a *high* magnetic field, decreasing it, lines 1 and 2 are swept first, yielding signals, and line 3 will then yield a signal too, but of opposite sign, since the populations of its corresponding levels are inverted. Thus, we obtain the signals indicated in Fig. 7.

Obviously, the signals do *not* possess the parity discussed in Section II. Note that an essential assumption for this argument was, that the intensities of lines 3 and 4 are considerably different. An actual calcula-

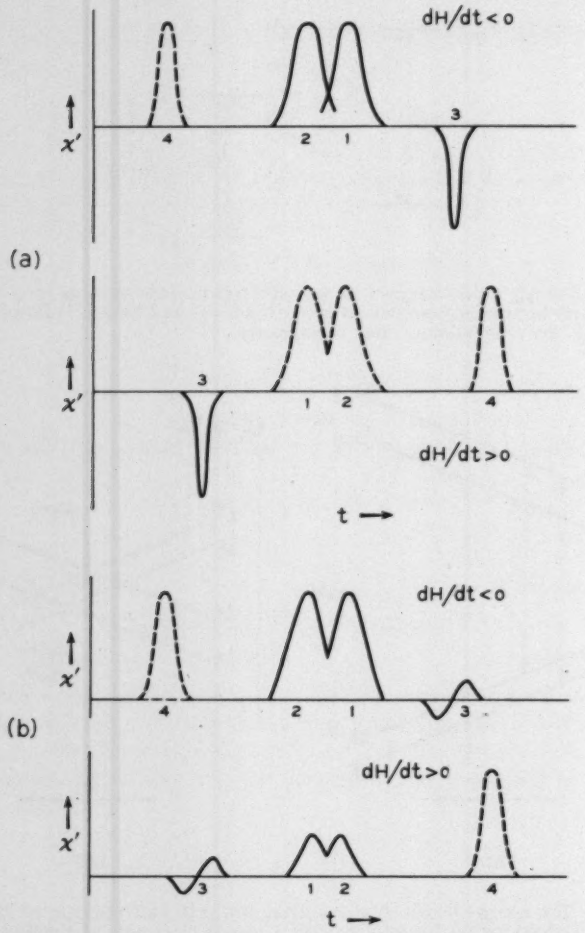


Fig. 7 — (a) The χ' signal expected from the system with energy levels shown in Fig. 5 when *all* transitions (including those of the forbidden lines), are adiabatic; line 4 is assumed to be *strictly* forbidden, while line 3 is not; there is *no* simple symmetry under $dH/dt \rightarrow -dH/dt$. (b) The χ' signal from the system with the energy level diagram shown in Fig. 5, when line 3 is *not* passed adiabatically.

tion indicates, however, that while the lines themselves are of order $[b/(\mu_n H)]^2$ weaker than the allowed ones, the difference between their intensities is of order $[ab/(\mu_e H)^2]^2$, i.e., $[a/(\mu_e H)]^2(\mu_n/\mu_e)^2$ times *less*. Therefore, for all practical purposes this asymmetry may be ignored.

Note that, for $b = 0$ and a microwave field *parallel* to the dc magnetic

field, line 3 is strictly forbidden, while line 4 is not. But line 4 is weak — of order $a^2/(\mu_e H)^2$ weaker than the allowed lines.

3.4 *Example 2*†

Consider a system of pairs of electrons resonating at slightly different frequencies $\omega \pm \Delta\omega$, due to different local fields, with an exchange interaction J between them. Let the eigenfunctions of each electron corresponding to the eigenvalues $m_s = +\frac{1}{2}$, $m_s = -\frac{1}{2}$, be $|\uparrow>$, $|\downarrow>$, respectively. Let $J > \Delta\omega$, $\gamma H_1 > \Delta\omega^2/J$. Then the eigenfunctions of the two-electron system are approximately $|\uparrow\uparrow>$, $|\downarrow\downarrow>$, $|\uparrow\downarrow> + |\downarrow\uparrow>$, $|\uparrow\downarrow> - |\downarrow\uparrow>$; and the energy levels of the system as functions of the magnetic field, in the fixed and rotating frames are shown in Fig. 8.

Assume the system to be at zero temperature. Then, if all passages are adiabatic, the χ' signal is shown in Fig. 9.

The central line will be seen only when we start from $H < H_0$ (or $H > H_0$, depending upon the sign of the exchange integral), with $H_0 = \omega/\gamma$.

Note the following points:

i. The "forbidden" lines at $\omega \pm J/\hbar$ are due to a mixture of the singlet and triplet $S_z = 0$ levels by the difference in local fields at the

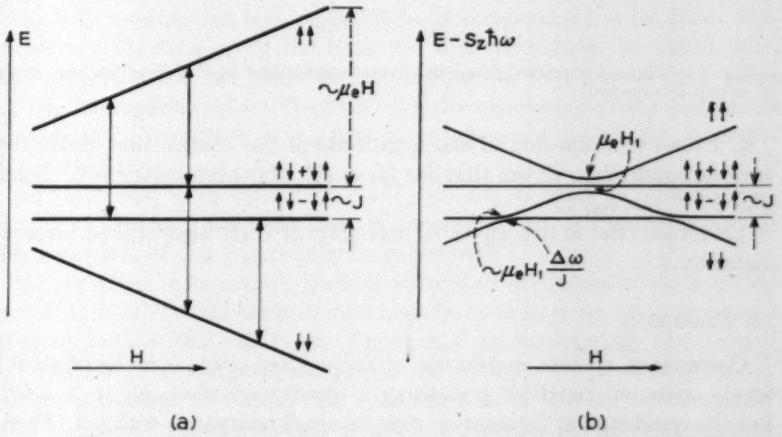


Fig. 8 — Energy levels of a system of two electrons, with a difference in hyperfine interactions of $\hbar\Delta\omega$, and an exchange integral J *strong* compared with the difference in hyperfine interactions: (a) energy levels in a fixed frame; (b) energy levels in a rotating frame. Again, an asymmetry in behavior under $dH/dt < 0$ and $dH/dt > 0$ is expected.

† This system has been treated in detail by P. W. Anderson (private communication).

two sites of the electrons. Therefore, their "intensity" will be reduced (relative to the "allowed" line) by about $(\hbar\Delta\omega/J)^2$. Thus, if $J \gg \hbar\omega$, they are not observed. A similar argument holds for $J \ll \hbar\Delta\omega$.

Thus, only electrons with an exchange interaction satisfying $J \approx \hbar\Delta\omega$ will give rise to the above phenomenon. The exchange interaction depends critically upon the distance between the electrons, thus only electrons at a "critical" distance apart may cause this asymmetry. The probability of finding electrons at this critical distance, in a physically diluted system, may be quite small.

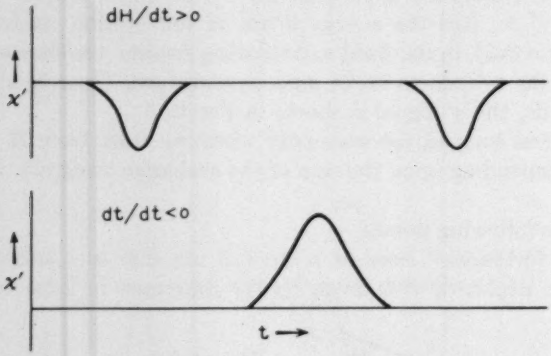


Fig. 9 — Signals expected from the system described in Fig. 8, when the sense of sweep is reversed.

ii. Closer examination of the amplitude of the central line under the various conditions, shows that for $\hbar\omega < kT$, the asymmetry is of order $(\hbar\omega/kT)^2$, not $\hbar\omega/kT$.

Therefore, the above effect is probably of little significance, experimentally.†

3.5 *Example 3*

Consider a system consisting of more than two (nonequidistant) levels, spin *one* particles possessing a quadrupole moment, in a solid. Let the quadrupolar interaction e^2qQ be small compared with μH. Then, the relative populations of the energy levels $S_z = -1$, $S_z = 0$, $S_z = 1$ are approximately $e^{\mu H/kT}$, 1 and $e^{-\mu H/kT}$, respectively. For *positive* eqQ and μ sweeping the magnetic field upward, the intensity of the first line

† However, this effect has been observed experimentally in a phosphorus doped silicon crystal, with impurity concentration of 6.5×10^{16} atoms/cm^3, at 1.2°K.

is proportional to $e^{\mu H/kT} - 1$, and that of the second† to $e^{\mu H/kT} - e^{-\mu H/kT}$, for adiabatic rapid passages neglecting relaxation.

Sweeping downwards, the intensities are $1 - e^{-\mu H/kT}$ and $e^{\mu H/kT} - e^{-\mu H/kT}$, respectively. For $\mu H/kT < 1$, the intensities are, approximately

$$\frac{\mu H}{kT} + \tfrac{1}{2}\left(\frac{\mu H}{kT}\right)^2, \qquad \frac{2\mu H}{kT} \qquad \text{going upwards}$$

and

$$\frac{\mu H}{kT} - \tfrac{1}{2}\left(\frac{\mu H}{kT}\right)^2, \qquad \frac{2\mu H}{kT} \qquad \text{going downwards}$$

indicating an asymmetry of order $(\mu H/kT)^2$.

From the above examples we may conclude that, although it may not be possible to prove the existence of a definite parity of ESR signals under the most general conditions, most physical systems seem to possess it, at least to a good approximation, even if they do not satisfy the the Bloch equations.

We also see the role played by the various conditions given by Bloch, as listed at the beginning of Section III:

i. The possibility of replacing all the interactions of a spin with its surroundings, by a "bath," is necessary, as shown by Examples 1 and 2.

ii. The assumption of a constant bath temperature is probably unnecessary. Heating up of the bath will tend to reduce the signal from those parts of the line swept later, compared to that due to those parts of the line swept earlier. This destroys the symmetry of the line about its center (if it existed in the first place), but not the parity discussed in Section II.

iii. We must have γH_1 , $1/T_1$, $1/T_2 \ll \gamma H_0$ in order to observe resonance at all (and, for wide lines, differences in the Boltzman factor at various parts of the line cannot be neglected).

iv. It seems to be always possible to expand the interactions in multipoles. If it is not, the spin system may be broken down to subsystems that do satisfy this condition. These may be interacting, but then i is violated. (For instance, the system of electrons with exchange interaction discussed in ii, may be considered as a system of "singlets" and "triplets," the interaction of which with their environment cannot be expanded in multipoles.)

v. Equidistant levels or a high temperature are necessary, as shown by Example 3.

† Since the population of the level $S_z = 0$ has been changed from "1" to $e^{\mu H/kT}$ by the adiabatic rapid passage through the line $S_z = 1 \leftrightarrow S_z = 0$.

IV. APPROXIMATE EXPRESSIONS FOR χ' AND χ'', AND FOR THE LOSS OF MAGNETIZATION FOR A SINGLE SPIN PACKET

Solutions of the Bloch equations for long relaxation times have been given by Jacobsohn and Wangness[3] and Salpeter.[4]

For an almost sudden transition $\gamma H_1^2 \ll dH/dt$ with constant dH/dt we have, according to Salpeter, for $dH/dt > 0$

$$\chi'(t) = \frac{H_0}{2H_1}\sqrt{\frac{2\gamma H_1^2}{dH/dt}}\,\chi_0\,\mathrm{Im}\left[e^{i\gamma(dH/dt)(t^2/2)}\int_{-\infty}^{\sqrt{\frac{\gamma(dH/dt)}{2}}t} e^{-iw^2}\,d\omega\right],$$

$$\chi''(t) = \frac{H_0}{2H_1}\sqrt{\frac{2\gamma H_1^2}{dH/dt}}\,\chi_0\,\mathrm{Re}\left[e^{i\gamma(dH/dt)(t^2/2)}\int_{-\infty}^{\sqrt{\frac{\gamma(dH/dt)}{2}}t} e^{-iw^2}\,d\omega\right]$$

For $dH/dt < 0$, χ' inverts sign and χ'' does not. It is assumed that, at $t = 0$, $H = H_0$ where $\gamma H_0 = \omega$. The loss of magnetization during an almost sudden transition is

$$\frac{\pi\gamma H_1^2}{|dH/dt|}\,|S|.$$

This solution, like the Bloch equations themselves, applies to a single spin packet (or, a homogeneously broadened line). This solution corresponds to the magnetization vector **S** moving on the surface of a sphere, starting from the "north pole," and eventually precessing around it in a circle of radius

$$\sqrt{\frac{2\pi\gamma H_1^2}{dH/dt}}\,|S|.$$

(See A in Fig. 10.) We have three physically distinct regions:

I. $\dfrac{1}{H_1}\dfrac{dH}{dt}t \ll 0$ ("before" resonance)†;

II. $\dfrac{1}{H_1}\dfrac{dH}{dt}t \approx 0$ ("on" resonance);

III. $\dfrac{1}{H_1}\dfrac{dH}{dt}t \gg 0$ ("after" resonance).

In region I, we have small χ' and χ'' signals, of the microwave frequency (approximately); χ'' is considerably smaller than χ'.

In region II, the signals are still of about the microwave frequency, and large; χ' and χ'' are of about the same strength.

† See Fig. 4 of Ref. 4.

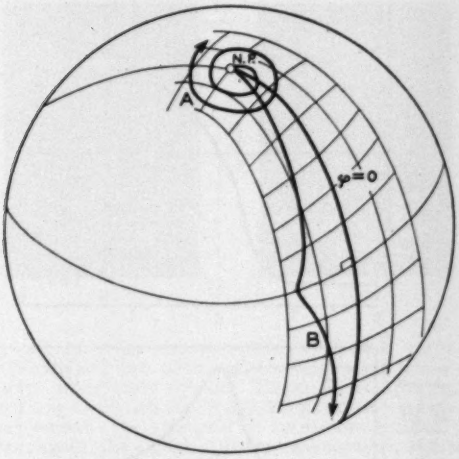


Fig. 10 — Locus of the magnetization vector for almost sudden (A) and almost adiabatic, (B), transitions for infinite relaxation times. (N.P. = North pole; φ = the polar angle.)

In region III, the signals are at a frequency γH, which is *not* the applied microwave frequency. These signals beat with the applied microwave frequency to produce "wiggles."

For an almost adiabatic transition ($\gamma H_1^2 > dH/dt$), we have (again, see Salpeter[4]):

$$\chi' = \pm \frac{1}{\sqrt{1 + \left(\frac{H - H_0}{H_1}\right)^2}} \chi_0 \frac{H_0}{2H_1},$$

$$\chi'' = + \frac{1}{\left[1 + \left(\frac{H - H_0}{H_1}\right)^2\right]^{\frac{1}{2}}} \frac{|dH/dt|}{\gamma H_1^2} \chi_0 \frac{H_0}{2H_1},$$

and the loss of magnetization due to nonadiabaticity is about (see Zener[13]):

$$\exp\left(-\frac{\pi \gamma H_1^2}{|dH/dt|}\right) |S|.$$

Again, this last solution applies to a single spin packet, or a homogeneously broadened line. The loss of magnetization at different values of $\gamma H_1^2/(dH/dt)$ is illustrated roughly in Fig. 11.

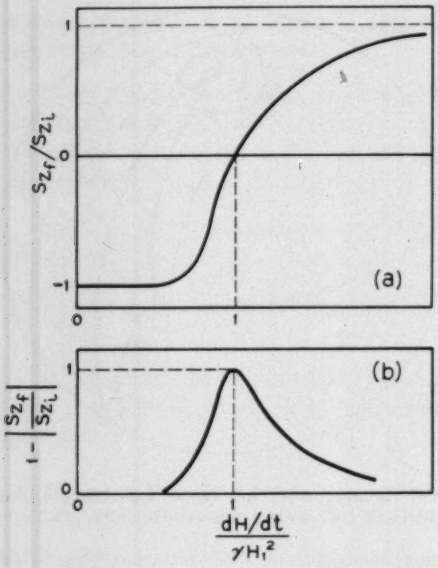


Fig. 11 — Loss of magnetization per passage due to nonadiabaticity (or nonsuddeness), qualitative only. In an adiabatic transition, $S_z \rightarrow -S_z$; in a sudden transition, $S_z \rightarrow S_z$; in both cases, there is no loss of $|\ S_z\ |$. In an intermediate transition, there is a loss in $|\ S_z\ |$; S_{z_i} and S_{z_f} are the initial and final values of S_z , respectively.

V. APPROXIMATE EXPRESSIONS FOR χ' AND χ'' FOR AN INHOMOGENEOUSLY BROADENED LINE

For an inhomogeneously broadened line for an almost sudden transition, under certain assumptions, χ' and χ'' are calculated in Appendix B. The results are

$$\chi'' = \frac{\pi}{2} H_0 \chi_0 h(H - H_0'),$$

$$\chi' = \tfrac{1}{2} H_0 \chi_0 \frac{dh(H - H_0')}{dH} \Delta H,$$

where $h(H - H_0')$ is the line shape [$\int h(H - H_0')dH = 1$]. These signals are the same as in the nonsaturated case [see Figs. 2(a) and 2(c), and also Fig. 12].

Physically, $S_z \approx S_0$ in *both* the nonsaturated and almost sudden cases. The difference between them is that, in the nonsaturated case,

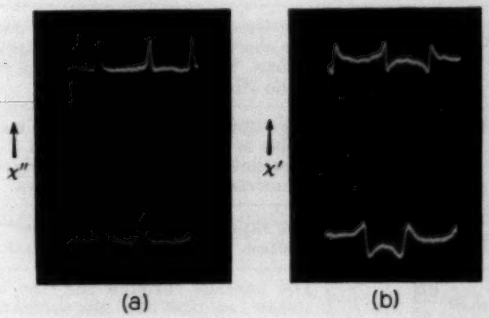


Fig. 12 — (a) χ'' signal of an almost sudden ($dH/dt \gg \gamma H_1^2$) transition of an inhomogeneously broadened line. Although the relaxation times are long, the line behaves as if it were completely relaxed. Data: $H_1 \approx 1/300$ gauss; $T = 1.2°K$ $T_1 \approx 300$ seconds); top trace, $dH/dt \cong 300{,}000$ gauss/second; bottom trace, $dH/dt \cong 100{,}000$ gauss/second. (b) χ' signal of an almost sudden transition, under the same conditions; again, the signal appears unsaturated, although $\gamma H_1 T_1 \approx 10^7$.

the definite phases relationships of the wave function (nondiagonal elements of the density matrix) are destroyed immediately by the relaxation process. In the "almost sudden" case, the relaxation times are long compared with the time in which the line is swept through; thus, these relaxation processes are less effective in destroying the phase relationships. However, the phases of different packets of the inhomogeneously broadened line, "cancel" each other, washing out the difference between it and the nonsaturated case. (However, see Appendix C.)

VI. THE VARIOUS PASSAGE CASES—DETAILED DISCUSSION

In this section, χ' and χ'' signals of an inhomogeneously broadened line will be discussed qualitatively and semiquantitatively. Various experimental conditions will be considered. The individual spin packets will be assumed to satisfy the Bloch equations approximately, and the loss of magnetization due to forbidden lines will be treated only phenomenologically.

The observed χ' and χ'' signals depend upon a variety of variables, which are listed at the end of this paper. To avoid lengthy repetitions, a shorthand notation will be used to describe some of the most frequently occurring relationships among the variables.

In this section, a magnetic field $H = H_0 + (dH_0/dt)t + H_m \cos \omega_m t$ will be postulated. The following terms signify the following relationships:

"Rapid": $\frac{H_1}{dH/dt} \ll \sqrt{T_1 T_2}$†	The time each spin packet is swept through, is *short* compared with the mean relaxation time. We have rapid passage conditions, in the Bloch[1] sense.
"Fast": $\omega_m T_1 \gg 1$	The time between successive modulation cycles is *short* compared with T_1. Each spin packet does not relax considerably between successive modulation cycles.
"Slow": $\frac{H_1}{dH/dt} \gg \sqrt{T_1 T_2}$	Not rapid, i.e., "slow passage" in the Bloch sense.
"Adiabatic": $\gamma H_1^2 \gg \frac{dH}{dt}$	The passage through the line is sufficiently slow, so that the magnetization vector **S** follows the effective magnetic field $\mathbf{H} - \boldsymbol{\omega}/\gamma + \mathbf{H}_1$ adiabatically (i.e., is parallel or antiparallel to it). Again, see Bloch.[1]
"Nonadiabatic": $\gamma H_1^2 \ll \frac{dH}{dt}$	The passage through the line is so rapid, that the magnetization vector **S** cannot follow the effective field, but stays almost parallel to the dc magnetic field.
"Burnt"	Most of the magnetization S has been destroyed ("burnt out") by the microwave field H_1.
"Stale"	Individual spin packets have reached a stationary state. The magnetization is an almost periodic function of time, with period $\omega_m/2\pi$.
"Fresh"	Most of the contribution to the signal comes from packets which have not yet been swept through many times.
"Scope Trace"	The signal at the output of the IF amplifier (or crystal detector, if no superheterodyne arrangement is used). This may be observed by an oscilloscope, or sometimes, a recorder.
"Recorder Trace"	The signal at the output of the phase sensitive detector.

† T_2 here, and in the following, is essentially T_{2e} of Redfield.[7] For perfect saturation it is approximately T_1. For imperfect saturation it is shorter. The situation $T_2 < H_1/(dH/dt) < \sqrt{T_1 T_2}$ corresponds to neither slow nor rapid passage. For $\gamma H_1 T_2 > 1$, we have $T_{2e} \approx T_1$, while for $\gamma H_1 T_2 < 1$, the signal is very weak (under saturation conditions). Therefore, this situation is ignored here.

Only rough estimates of the amplitudes are given, mainly in order to indicate qualitatively the expected effect of variation of the experimental parameters. In general, it will be assumed that $H_1 < H_m$. This restriction, however, is not very severe. If it does not hold, the logarithmic dependence $\log (H_m/H_1)$ (if any) is replaced by a linear dependence H_m/H_1.

The discussion of this section is rather tedious; therefore, the results are summarized in Section VII.

6.1 *Case 1 — Slow Passage*

Physical Description

The time of passage through each individual spin packet, $t \approx H_1/\omega_m H_m$, is long compared with $\sqrt{T_1T_2}$. The system is in equilibrium at any moment.

This case has been treated by Portis,[5] ("Case I") and Portis.[11]

Conditions Under Which This Case Is Observed:

$$\frac{1}{\gamma H_1} \ll \sqrt{T_1T_2} \ll \frac{H_1}{\omega_m H_m}; \qquad H_m \ll \Delta H; \qquad \sqrt{T_1T_2} \ll \frac{H_1}{\omega_m \Delta H}.$$

Note that, for given $\omega_m \Delta H$ and a bounded H_1, this case may apply only for a narrow range of T_1 (if at all). Wider lines, and high modulation frequencies, are unfavorable for this case. [The condition $\sqrt{T_1T_2} \ll H_1/\omega_m \Delta H$ follows, from comparison of the amplitudes in this case, with those of case 2 in Section 6.2. If this condition does not hold, the signal due to the second case predominates.]

Traces and Amplitudes

For slow passage, the solution of the Bloch equation for a single spin packet is[1]

$$\chi' = \tfrac{1}{2}\chi_0\omega T_2 \frac{(\omega - \omega_0)T_2}{1 + (\gamma H_1)^2 T_1T_2 + (\omega - \omega_0)^2 T_2^2},$$

$$\chi'' = \tfrac{1}{2}\chi_0\omega T_2 \frac{1}{1 + (\gamma H_1)^2 T_1T_2 + (\omega - \omega_0)^2 T_2^2}.$$

Under saturation, $(\gamma H_1)^2 T_1T_2 \gg 1$. Furthermore, $\omega - \omega_0 \ll \omega$. Thus

$$\chi' \cong \tfrac{1}{2}\chi_0\omega \frac{\omega - \omega_0}{(\gamma H_1)^2 \dfrac{T_1}{T_2} + (\omega - \omega_0)^2},$$

$$\chi'' \cong \tfrac{1}{2}\chi_0 \frac{\omega}{(\gamma H_1)^2 T_1 + (\omega - \omega_0)^2 T_2}.$$

Now, consider an inhomogeneously broadened line, of shape $h(H - H_0')$ and width ΔH. If $H_1\sqrt{T_1/T_2} \ll \Delta H$, for most spin packets $(\omega - \omega_0)^2 \gg$

$(\gamma H_1)^2(T_1/T_2)$, and the packets for which $(\omega - \omega_0)^2 < \gamma^2 H_1^2(T_1/T_2)$, will give almost equal and opposite contributions to χ'. Then, for the whole line

$$\chi'(H) \approx \tfrac{1}{2}\chi_0 H_0 \int \frac{H - H_0}{H_1^2 \dfrac{T_1}{T_2} + (H - H_0)^2} h(H_0 - H_0')\, dH_0$$

$$\approx \tfrac{1}{2}\chi_0 H_0 \int \frac{1}{(H - H_0)} h(H_0 - H_0')\, dH_0$$

$$= -\chi_0 \int_0^\infty H h(H_0 - H_0') \frac{H_0\, dH_0}{H_0^2 - H^2}.$$

This expression has been given by Portis.[5] Note that the integral does not converge, and must be understood in the sense of its principal part. [The packets for which $(H - H_0)^2 < H_1^2(T_1/T_2)$ will tend to give equal and opposite contributions, which cancel.] For χ'', we cannot make this approximation, because the contributions from $(H - H_0)^2 < H_1^2(T_1/T_2)$ do not cancel, but add up. However, if $\Delta H^2 \gg (\gamma H_1)^2(T_1/T_2)$, $h(H_0 - H_0')$ can be assumed constant and withdrawn outside the integral (for χ' this approximation yields zero, and thus cannot be applied):

$$\chi'' = \frac{\tfrac{1}{2}\chi_0 H_0}{\gamma} \int \frac{h(H_0 - H_0')\, dH_0}{H_1^2 T_1 + (H_0 - H)^2 T_2}$$

$$\approx \frac{\tfrac{1}{2}\chi_0 H_0 h(H_0 - H_0')}{\gamma} \int \frac{dH_0}{H_1^2 T_1 + (H_0 - H)^2 T_2}$$

$$= \frac{\pi}{2} \chi_0 \omega \frac{1}{(\gamma H_1)\sqrt{T_1 T_2}} h(H_0 - H_0')$$

$$= \frac{\pi}{2} \chi_0 \frac{H_0}{H_1} \sqrt{\frac{1}{T_1 T_2}} \frac{h(H_0 - H_0')}{\gamma}.$$

When the field is modulated, $\Delta\chi' = (d\chi'/dH_0)H_m \cos \omega_m t$ for each packet, and signals of different packets must be superimposed:

$$\int \frac{d\chi'(H - H_0')}{dH} h(H_0' - H_0)\, dH_0$$

$$= \frac{d}{dH} \int \chi'(H - H_0) h(H_0' - H_0)\, dH_0$$

$$= \frac{d\chi'(H, H_0')}{dH}.$$

Thus, the recorder trace is proportional to the derivative of the line shape. For

$$h(H_0' - H_0) = \frac{1}{\sqrt{2\pi}\Delta H} e^{-(H_0'-H_0)^2/2\Delta H^2}$$

at the center of the line, the amplitude is

$$-\left(\frac{d\chi'}{dH}\right)_{H=H_0'} \approx \tfrac{1}{2}\chi_0 H \int \frac{1}{H - H_0} \frac{dh(H_0' - H_0)}{dH'} dH_0$$

$$= \tfrac{1}{2}\chi_0 \frac{H}{\Delta H^2}.$$

Thus, the maximum value of the fundamental frequency component of χ', denoted χ_f', is

$$\chi_f' = \frac{1}{2}\left(\frac{H}{\Delta H}\frac{H_m}{\Delta H}\chi_0\right).$$

The same argument also holds for χ'', and we have

$$\chi_f'' = \frac{1}{2}\sqrt{\frac{\pi}{2}}\left(\frac{H_0}{\Delta H}\frac{H_m}{\Delta H}\chi_0\right)\frac{1}{\gamma H_1\sqrt{T_1T_2}}.$$

Experimental Data

A considerable amount of quantitative experimental data applying to this case has been given by Portis.[11] It is not felt necessary to repeat similar data here.

6.2 *Case 2 — Rapid Adiabatic Passage with a Long Time Between Consecutive Field Modulation Cycles*

Physical Description

The time of passage through each individual spin packet is small compared with $\sqrt{T_1T_2}$. Thus, each packet yields an adiabatic rapid passage signal. The time between consecutive sweeps of each packet is long compared with T_1, so that each packet has sufficient time to relax completely.

Conditions Under Which This Case Is Observed:

$$\frac{H_1}{\omega_m H_m} \ll \sqrt{T_1T_2}, \quad \omega_m T_1 \ll 1, \quad \gamma H_1^2 \gg \omega_m H_m, \quad \gamma H_1\sqrt{T_1T_2} \gg 1.$$

Traces

The χ' signal is positive for $dH/dt < 0$, negative for $dH/dt > 0$. Thus, we have the trace shown in Fig. 13 (χ' is practically independent of the rate of change of the magnetic field). The output of the phase-sensitive detector will be proportional to $h(H_0 - H_0')$ [Fig. 3(d)].

The χ'' signal is negligible. If it is present, it may be due to incomplete saturation, partial breakdown of the adiabatic condition or slow passage near the extrema of the modulation cycle. In both cases, the recorder trace will be the derivative of the line shape (since χ'' due to slow passage, or to a breakdown of the adiabatic condition, is inherently of a frequency *twice* the field modulation frequency, and thus not detected by the phase sensitive detector. Only the derivative of the line shape yields a nonvanishing result).

Amplitudes

For each individual spin packet, χ' is proportional to

$$\frac{1}{\sqrt{H_1^2 + (H - H_0)^2}}.$$

For an infinitely wide line,

$$\chi' \propto \int_0^\infty \frac{dH_0}{\sqrt{H_1^2 + (H - H_0)^2}},$$

which is infinite. Therefore, to obtain a finite result, the finite width of

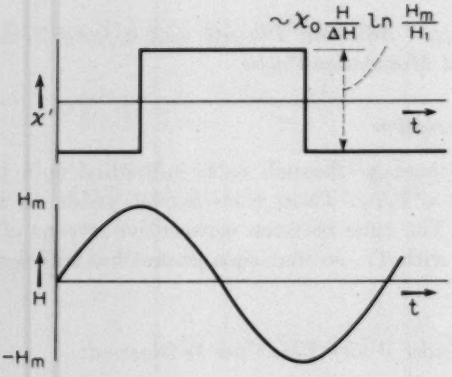


Fig. 13 — Scope trace of χ' signal to be expected in case 2. Ideal conditions are assumed; the field modulation amplitude H_m is small compared with line width ΔH.

the spin packet distribution has to be called upon. This finite width is due to the finite width of the inhomogeneously broadened line itself, and the reduction of the contribution of packets far off resonance due to relaxation. This reduction is difficult to calculate exactly. However, the integral in the expression for χ' above diverges only logarithmically; therefore, variations in the integrand far away from resonance, effect the integral only little. An approximate calculation has been performed by Portis,[5] and his result is

$$\chi_f' = \chi_0 H_0' h(H - H_0') \ln \frac{2H_m\omega_m T_1}{H_1}$$

for $\omega_m H_m \sqrt{T_1 T_2} \gg H_1$ and $T_1 = T_2$. [Portis, Case II(b).] This expression is discussed in Appendix D.

The χ'' amplitude is, for an individual spin packet (for χ'' due to breakdown of the adiabatic condition; see Section IV):

$$\chi'' = \tfrac{1}{2}\chi_0 \frac{H_0}{H_1} \frac{|\gamma H/dt|}{\gamma H_1^2} \frac{1}{\left[1 + \left(\frac{H - H_0}{H_1}\right)^2\right]^{\frac{1}{2}}}.$$

Now,

$$\int_{-\infty}^{\infty} \frac{dH}{\left[1 + \left(\frac{H - H_0}{H_1}\right)^2\right]^{\frac{1}{2}}} = 2H_1.$$

Thus, superimposing various packets:

$$\chi'' = \chi_0 H_0 \frac{|dH/dt|}{\gamma H_1^2} h(H - H_0') \approx \frac{2}{\pi} \chi_0 H_0 \frac{H_m\omega_m}{\gamma H_1^2} h(H - H_0')$$

$$+ \frac{2}{\pi} \chi_0 H_0 \frac{H_m\omega_m}{\gamma H_1^2} h(H - H_0') \cos 2\omega_m t$$

$$+ \frac{2}{\pi} \chi_0 H_0 \frac{H_m\omega_m}{\gamma H_1^2} \frac{dh(H - H_0')}{dH} H_m \cos \omega_m t.$$

Since the χ'' lines due to an individual spin packet, caused by the breakdown of the adiabatic condition are narrow, neglect of relaxation is justified

Comments

For a sinusoidal field modulation, the passage must be rapid throughout the field sweep and not just in the middle of it. The time between

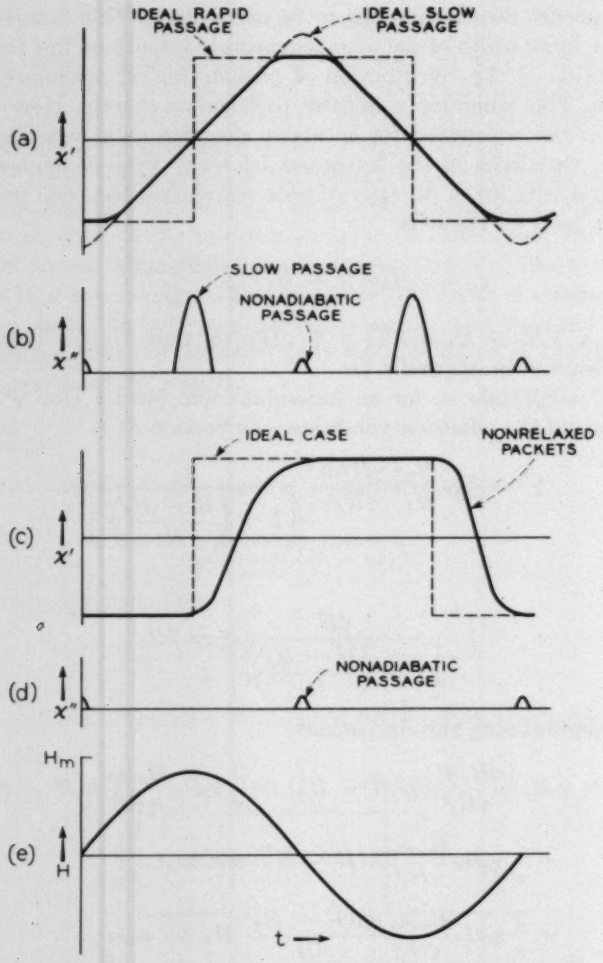


Fig. 14 — Scope traces of χ' signal to be expected when the conditions are not ideal: (a) χ' signal when there is *slow* passage near the extrema of the modulation cycle, where dH/dt is small; (b) χ'' signal when there is slow passage near the extrema of the modulation cycle; there may also be some signal due to imperfect adiabatic conditions, when $| dH/dt |$ is large; (c) χ' signal when T_1 is not very short compared with the period of the modulating field; spin packets near the extrema of the modulation cycle do not have time to relax completely before being swept over again; (d) χ'' signal under the conditions of (c); (e) the magnetic field modulation cycle.

consecutive sweeps of the same package, must be large compared with T_1 near the extrema of the modulation cycle, too.

These limitations are quite severe. Very often, T_1 will be a little short, and then the passage near the extrema will be slow [Figs. 14(a), (b)], or T_1 will be a little too long, and then the packages near the end will not have sufficient time to relax completely [Figs. 14(c), (d)].

Experimental Data

(*a*) *Fig. 15.*

Data: χ' trace; $H_1 \approx 1/30$ gauss; $H_m = 1$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 10°$K. Under the same conditions, the χ'' trace was small.

Conditions that are satisfied: H_1, $H_m < \Delta H$; $\gamma H_1 T_1 > 1$; $\gamma H_1^2 > \omega_m H_m$; $\omega_m T_1 < 1$; ($T_2 = T_1 \approx 10^{-5}$ second; $\gamma H_1 T_1 \approx 5$; $\gamma H_1^2 \approx 15000$; $\omega_m H_m \approx 6000$; $\omega_m T_1 \approx 1/20$).

Interpretation: $H_1/(\omega_m H_m) \approx \frac{1}{2} \times 10^{-5}$ second. This is of order T_1. The passage in the *middle* of the modulation cycle is rapid, while at the extrema it is slow. The signal was observed to be just a little less than 90° out-of-phase with the field modulation. This indicates that this is *not* case 1, but rather a (distorted) case 2 [Portis,[5] Case II(b)]. The extrema just commence to flatten [see Fig. 14(a)].

(*b*) *Fig. 16 — Top and Center Traces*

Data: $H_1 \approx 1/100$ guass; $H_m = 1/10$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 10°$K (top = χ', center = χ'').

Conditions that are satisfied: H_1, $H_m < \Delta H$; $\gamma H_1 T_1 \approx 1$; $\gamma H_1^2 > \omega_m H_m$; $H_1/\omega_m H_m \approx T_1$; ($\gamma H_1^2 \approx 1500$; $\omega_m H_m \approx 600$; $\gamma H_1 T_1 \approx 2$; $H_1/\omega_m H_m \approx T_1$); $T_1 = T_2$.

Interpretation: Here, $\gamma H_1 T_1$ was reduced so that a χ'' signal can be observed at a reasonable amplitude. This, however, causes the conditions of case 2 not to hold very well, and the χ' trace is a little distorted.

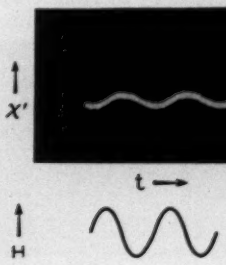


Fig. 15 — Experimental scope trace under the conditions of Fig. 14(a).

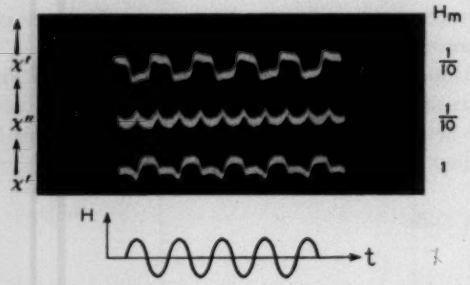


Fig. 16 — Experimental scope traces for case 2. Top trace, χ' signal, almost ideal case "2"; center trace, χ'' signal due to slow passage near the extrema of the modulation cycle; bottom trace, χ' signal, when the adiabatic condition is slightly violated.

(c) *Fig. 17.*

Data: χ'; $H_1 \approx 1/100$ gauss; $dH/dt = 1/10$ gauss/second; over-all sweep: 1.6 gauss; $T = 4.5°\text{K}$; $T_1 \approx 5$ seconds.

Conditions that are satisfied: "H_m", $H_1 < \Delta H$; $\gamma H_1 \sqrt{T_1 T_2} \gg 1$; $\gamma H_1^2 \gg dH/dt$; $H_1/(dH/dt) > \sqrt{T_1 T_2}$.

Interpretation: This is the χ' signal observed at the output of the IF amplifier (no phase sensitive detector). It corresponds to case 2, with a slight amount of case 7 (i.e., incomplete relaxation between successive field sweeps) intermixed [see Fig. 14(c)].

6.3 *Case 3 — Rapid Passage with a Long Time Between Consecutive Field Modulation Cycles, the Adiabatic Condition Being Slightly Violated*

Physical Description

The conditions under which case 2 holds are quite severe, and it may be quite often impossible to satisfy them with the experimental equipment at hand, and it will be necessary to violate some of these conditions.

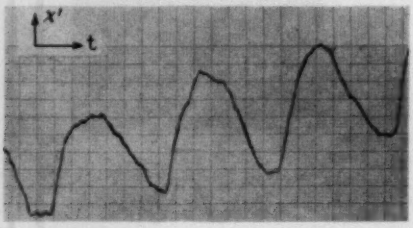


Fig. 17 — χ' signal for rapid adiabatic passage, when T_1 is not very small compared with the field modulation cycle. Compare with Fig. 14(c). No phase sensitive detector is used here (therefore, considerable drift is observed).

If the condition $\omega_m H_m < \gamma H_1^2$ is violated, the adiabatic condition will break down near the centers of the modulation cycle. (This case is not of much theoretical significance, but it is experimentally very common.)

The modulation cycle may be subdivided into a number of regions (see Fig. 18):

Region I: $H_1/(dH/dt) > \sqrt{T_1 T_2}$. In this region we have slow passage conditions.

Region II: $1/(\gamma H_1) < H_1/(dH/dt) < \sqrt{T_1 T_2}$. In this region we have adiabatic rapid passage conditions.

Region III: $H_1/(dH/dt) < 1/(\gamma H_1)$. In this region the adiabatic condition is broken.

We have the further possibilities:

i. $T_1 < \sqrt{H_1/(\omega_m^2 H_m)}$. In this case, $t = T_1$ lies in region I.

ii. $\sqrt{H_1/(\omega_m^2 H_m)} < T_1 < \gamma H_1^2/(\omega_m^2 H_m)$. In this case, $t = T_1$ lies in region II.

iii. $\gamma H_1^2/(\omega_m^2 H_m) < T_1$. In this case, $t = T_1$ lies in region III.

Case i: In region I we have slow passage; in region II we have adiabatic rapid passage of fully relaxed packets; and in region III nonadiabatic passage of fully relaxed packets.

Case ii: Even in region II there are some packets which did not have time to relax since last swept by the field. If $T_1 = T_2$, region I is then, naturally, negligibly small (less than one packet wide). In region III, all packets have fully relaxed.

Case iii: The packets in region II have little time to relax. Even some of the packets in region III did not relax.

Generalizing to systems with $T_2 < T_1$ will create additional cases, since one has to distinguish between relaxation of S_z, and of S_x, S_y.

Traces

See Figs. 18 and 19. In the traces of cases ii and iii, region I has been omitted, since it is small. The traces have been drawn for the center of the line. Off center, the derivative of the line shape will yield an additional χ' signal. In addition, there may be "wiggles" in the border between regions II and III.

The traces detected by a phase-sensitive detector are "absorption" curves, at a phase angle of 90° with the modulating field in case i, 0° in case iii, and an intermediate value for case ii. In addition, the χ' signal of region III will tend to produce a "dispersion derivative." (The signals of regions I and II also produce such a signal, but this is weak compared with the "absorption" shaped signal. If region III is *large*, its "dispersion derivative" signal may become comparable, or exceed the

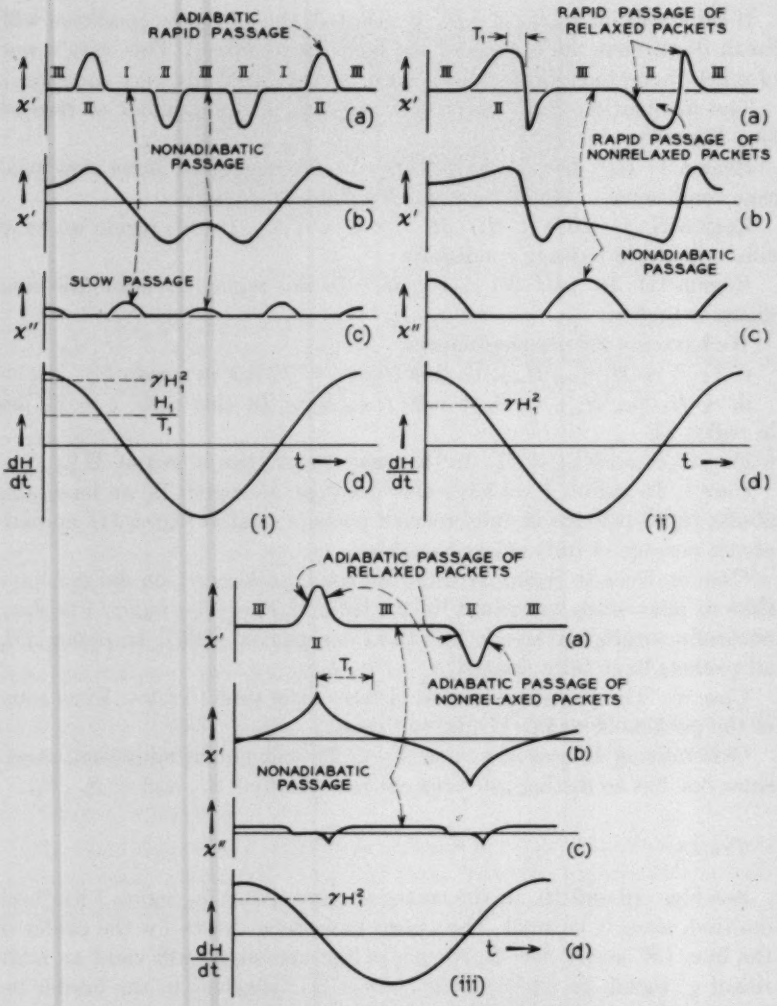


Fig. 18 — Signal to be expected when the adiabatic condition $\gamma H_1^2 \gg dH/dt$ is broken: (i) Short T_1 ; there is slow passage near the extrema of the field modulation cycle, and the spin packets are completely relaxed in the other regions. (ii) Slightly longer T_1 ; some spin packets which are swept through adiabatically, are not relaxed. (iii) Still longer T_1 ; all spin packets which are swept through adiabatically, are not relaxed. In all three sections, (a) is an "idealized" χ' trace, (b) is a more "realistic" χ' trace, (c) is the χ'' trace and (d) is the magnetic field modulation.

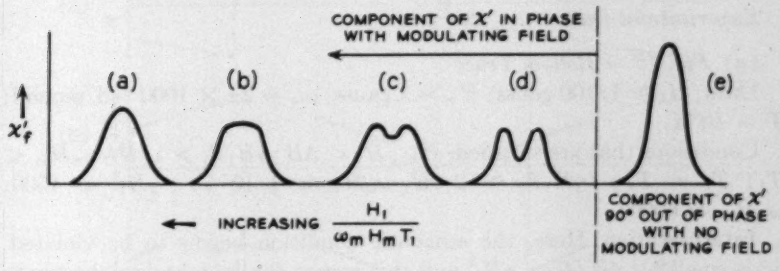


Fig. 19 — Expected recorder traces for rapid (adiabatic or nonadiabatic) passage, with a field modulation period long compared with, or of the same order, as T_1 : (a) large H_1 , small H_m ; (b) and (c) intermediate values of H_1 , H_m ; (d) Small H_1 , large H_m . (These four are the component of χ_f' in-phase with the field modulation.) (e) Component of χ_f' 90° out-of-phase with the field modulation.

"absorption" signal due to region II.) The "dispersion derivative" signal is at 0° with respect to the modulating field, so we may expect χ' traces as shown in Fig. 19.

Note that region III also yields a χ'' signal component at the modulation frequency, proportional to dh/dH; this component (which yields an "absorption derivative" trace) is small.

Amplitudes

I shall not attempt a quantitative treatment of this case. However, if the relative depth of the "notches" of cases i and ii can be measured, and is found to be β, the fundamental frequency component falls down by $(2/\pi)\beta$, for the simplified shape function of Fig. 20.

Case iii is essentially identical with case 10.

Making use of the formula for χ' of case 2:

$$\chi_f' \approx \left(\frac{2}{\pi}\beta\right) \chi_0 H h (H - H_0') \ln \frac{2H_m\omega_m T_1}{H_1}.$$

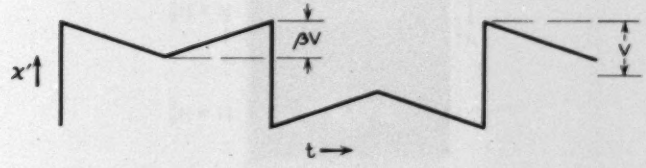


Fig. 20. — An idealized χ' scope trace of case 3, used to estimate the amplitude of the signal.

Experimental Data

(*a*) *Fig. 16 — Bottom Trace*

Data: $H_1 \approx 1/100$ gauss; $H_m = 1$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 10°K$.

Conditions that are satisfied: H_m , $H_1 < \Delta H$; $\gamma H_1 T_1 > 1$; $H_1/\omega_m H_m < T_1$; $T_1 = T_2$; ($\gamma H_1 T_1 \approx 2$; $H_1/\omega_m H_m \approx 1/10\ T_1$; $\gamma H_1^2 \approx 1500$; $\omega_m H_m \approx 6000$).

Interpretation: Here, the adiabatic condition begins to be violated. χ' is small for $dH/dt > \gamma H_1^2$ and this causes the 'notches' in the photo [see (i)-(b) in Fig. 18].

The observed trace probably includes a small χ'' component [see (i)-(c) in Fig. 18], which adds up with χ' to yield the observed trace.

(*b*) *Fig. 21.*

Data: χ'; $H_1 \approx 1/100$ gauss; $H_m = 1$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 10°K$ [same as in (a) above, except for higher scope gain]; bottom trace: sitting in the center of the line; top and center traces: off the center of the line.

Conditions: the same as in (a).

Interpretation: Since the condition $H_m \ll \Delta H$ is not very well satisfied, a slight admixture proportional to the derivative of the line shape, distorts the shape somewhat. This is particularly evident for the traces off the center of the line.

Synthesis of the observed traces from those due to various cases is shown in Fig. 22.

(*c*) *Fig. 23.*

Data: χ''; $H_1 \approx 1/100$ gauss; $H_m = 1$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 10°K$; top and center traces: off the center of the line;

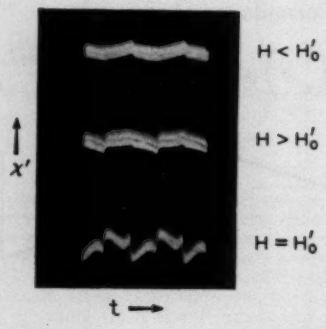


Fig. 21 — χ' scope traces of rapid, nonadiabatic passage of a line, when T_1 is very short compared with the field modulation cycle, at the center of the line (bottom trace), and on both sides of it (top and center traces).

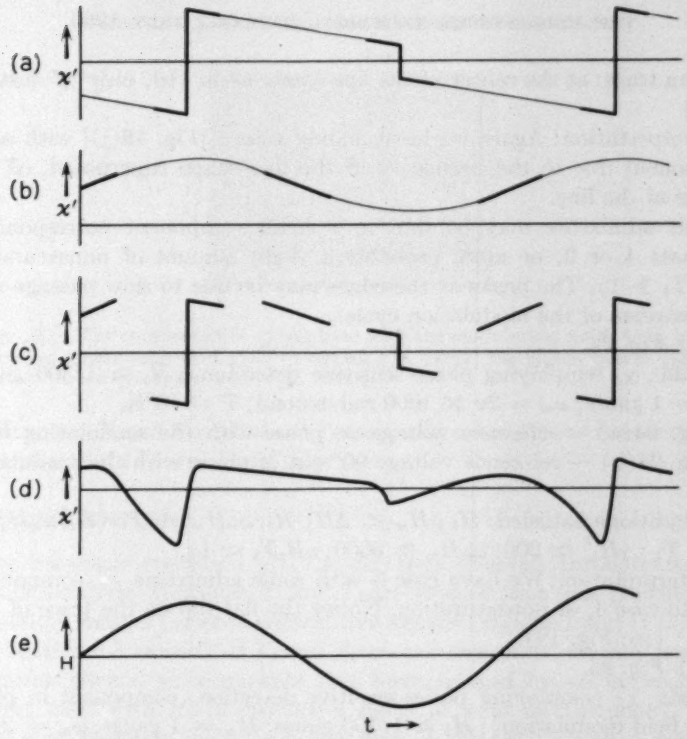


Fig. 22 — Synthesis of the χ' trace observed in Fig. 21: (a) a case 2 signal (adiabatic rapid passage, slow modulation), on the *shoulders* of the line; (b) a case 4 signal (nonadiabatic); (c) synthesis of (a) for the portions of the modulation cycle for which dH/dt is small, and (b) for those for which dH/dt is large; (d) rough shape of the trace expected experimentally; (e) the modulating field.

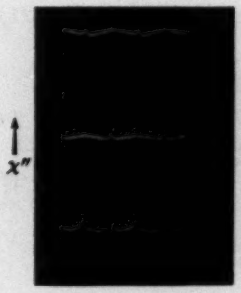


Fig. 23 — χ'' scope traces under the same conditions as those of Fig. 21. We observe slow passage signals, with some component proportional to the derivative of the line shape, due probably to incomplete saturation. ($\gamma H_1 T_1 \approx 2$.)

bottom trace: at the center of the line [*same* as in (b), only χ'' instead of χ']

Interpretation: Again we have mainly case 3 [Fig. 18(i)] with some component due to the derivative of the line shape superposed, off the center of the line.

This admixture may be due to a small component corresponding to cases 4 or 9, or more probably a slight amount of nonsaturation ($\gamma H_1 T_1 \not> 1$). The peaks at the edges may be due to slow passage near the extrema of the modulation cycle.

(*d*) *Fig. 24.*

Data: χ_f' (employing phase-sensitive detection); $H_1 \approx 1/300$ gauss; $H_m = 1$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 10°$K.

Fig. 24(a) — reference voltage *in phase* with the modulating field;

Fig. 24(b) — reference voltage 90° out of phase with the modulating field.

Conditions satisfied: H_1, $H_m < \Delta H$; $H_1/\omega_m H_m < T_1$ ($H_1/\omega_m H_m \approx 1/30\ T_1$; $\gamma H_1^2 \approx 200$; $\omega_m H_m \approx 6000$; $\gamma H_1 T_1 \approx 1$).

Interpretation: We have case 3 with some admixture of components due to case 4, or nonsaturation. Notice the flat top on the trace of Fig. 24(a) (see Fig. 19).

(*e*) *Fig. 25.*

Data: χ_f' (employing phase-sensitive detection; component in phase with field modulation); $H_1 \approx 1/100$ gauss; $H_m = 1$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 10°$K.

Conditions satisfied: H_1, $H_m < \Delta H$; $\gamma H_1 T_1 > 1$; $H_1/\omega_m H_m < T_1$; $T_1 = T_2$ ($\gamma H_1 T_1 \approx 2$)

Interpretation: The increased saturation, and more important, the less violent violation of the adiabatic condition, compared with Fig. 24(a), gives a "cleaner" trace (χ', without phase-sensitive detection under the same conditions, is shown in Fig. 21).

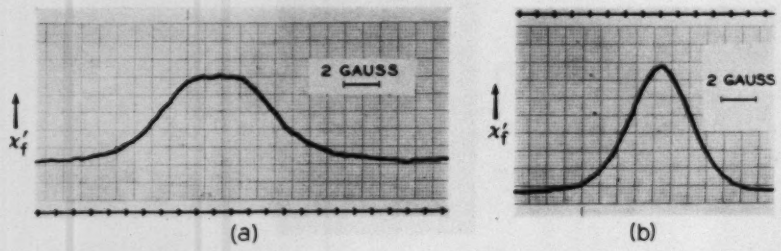


Fig. 24 — (a) χ' for mixture of cases 1 and 2; with P.S.D., 0° phase shift between modulating field and reference voltage [compare with Fig. 19(b)]. (b) Same as (a), but χ' signal 90° out-of-phase with modulating field.

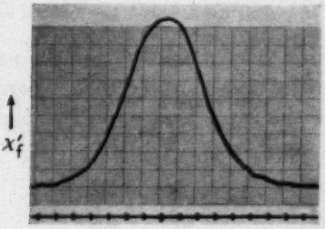


Fig. 25 — The component of χ_f' in-phase with the modulating field, with higher microwave power than that used to obtain Fig. 24(a); the higher power yields a "cleaner" trace [compare with Fig. 19(a)].

6.4 *Case 4 — Rapid Nonadiabatic Passage with a Long Time Between Consecutive Field Modulation Cycles*

Physical Description

The adiabatic condition $\gamma H_1^2 > dH/dt$ is violated throughout the modulation cycle. However, T_1 and T_2 are so short that spin packets are completely relaxed between consecutive sweeps (regions I and II of case iii in Section 6.3 vanish). If T_1 and T_2 are not *very* short, we may neglect relaxation effects, since packets that have relaxed by an appreciable amount, beat with H_1 to produce a signal of a far off frequency. The signal will, therefore, be a "superposition of wiggles," which is not detected by the narrow bandwidth system usually employed, as treated in Section V. (Even if T_1 , T_2 are shorter, their effect will probably not be large. The effect of T_2 is to destroy the phases — but phases of various packets, interfere destructively, as discussed in Section V; T_1 increases S_z , but, for $\gamma H_1^2 \ll dH/dt$, S_z does not fall off much, anyway.)

Conditions Under Which This Case Is Observed:

$$\omega_m T_1 < 1; \quad \gamma H_1^2 \lll H_m \omega_m ; \quad H_m < \Delta H; \quad \gamma H_1 \sqrt{T_1 T_2} \gg 1.$$

Note that the condition $\gamma H_1^2 \ll H_m \omega_m$ requires a very small H_1 , while saturation requires that it be not too small. Thus, this case may be difficult to achieve practically. A wider line makes it possible to increase H_m , and thus, makes the observation of this case easier.

Traces

For oscilloscope traces, see Fig. 26.

The output of the phase-sensitive detector will consist of a "dispersion derivative" line for χ' and a "absorption derivative" for χ''.

Amplitudes

As discussed in Section 3.5, χ' and χ'' for the extreme nonadiabatic case are the *same* as for the nonsaturated case. Therefore, the line shapes and amplitudes will be the same here as in the nonsaturated case, (although T_1, T_2 may be quite large):

$$\chi_f' \approx \tfrac{1}{2} \frac{H}{\Delta H} \chi_0 \frac{H_m}{\Delta H},$$

$$\chi_f'' \approx \tfrac{1}{2} \frac{H}{\Delta H} \chi_0 \frac{H_m}{\Delta H}.$$

6.5 *Case 5 — Rapid Adiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles and Very Rapid Field Sweep*

Physical Description

In this case, each spin package yields an adiabatic rapid passage signal. The time between consecutive modulation cycles is small compared with T_1. In addition, the conditions are such that the packages do not lose much of their magnetization during the whole process. [$H = H_0 + (dH_0/dt)\, t + H_m \cos \omega_m t$. If dH_0/dt is sufficiently large, this may be the case.]

This case has been treated by Portis,[5] and called "Case IV" by him.

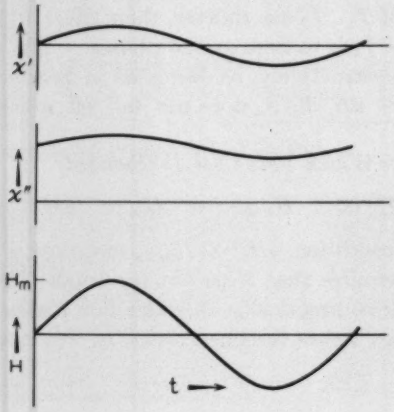


Fig. 26 — Expected scope trace of signal of case 4 (spin packets are completely relaxed; χ'' signal *off* the center of the line).

Conditions Under Which This Case Applies:

$$\omega_m T_1 \gg 1; \qquad \gamma H_1^2 \gg H_m \omega_m ; \qquad \gamma H_1 \sqrt{T_1 T_2} \gg 1.$$

In addition, assume that a fraction α of the line is "burnt" during each passage. Then, the average of the magnetization, (and signals) of N passages will be

$$\frac{1}{N}[1 + (1 - \alpha) + (1 - \alpha)^2 + \cdots] = \frac{1 - (1 - \alpha)^N}{\alpha N} \approx \frac{1 - e^{-\alpha N}}{\alpha N}.$$

The number of times each packet is swept over is

$$N = \frac{2H_m}{dH_0/dt}\left(2\frac{\omega_m}{2\pi}\right).$$

Thus, for little burnout, we must have; $\alpha N \ll 1$, or $[H_m \omega_m/(dH/dt)]\alpha \ll 1$.

The amplitude of the signal in this case will be about $H_m/\Delta H$ weaker than in case 6. Therefore, the above condition is a little too soft, and we must have

$$\frac{H_m \omega_m}{dH_0/dt}\alpha \ll \frac{H_m}{\Delta H}, \qquad \frac{\Delta H \omega_m}{dH_0/dt}\alpha \ll 1.$$

The loss of magnetization during rapid passage, α, also depends very much upon the sample used. If the loss is mainly due to relaxation between consecutive modulation cycles, $\alpha \approx 1/(\omega_m T_1)$, as postulated by Portis.[5]†

Traces

Each spin packet yields the traces shown in Fig. 27 (χ'' is due to nonperfect adiabaticity). Notice that the various packets will combine to produce a χ' signal that is approximately constant and a sinusoidal χ'' signal ($\chi'' \propto dH/dt \approx \omega_m H_m \sin \omega_m t$ for each packet). The derivative of the line shape will yield variations in the χ' signal, and thus a component of the signal at the modulation frequency (Fig. 28). Thus the recorder traces are as shown in Fig. 29. The χ' signal is of "absorption derivative" shape, and χ'' of "absorption" shape. However, the parities are as discussed in Section II.

Amplitudes

The amplitude of the χ' trace depends quite critically upon α, which is a rather indefinite quantity. It depends upon the rate at which a spin

† For the sample employed in this experiment, the loss was found to be due mainly to nonadiabatic passage, and to forbidden lines (Section 3.3).

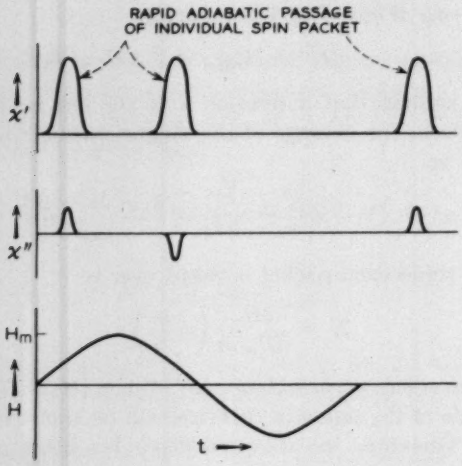


Fig. 27 — Signals from one spin packet, case 5; the spins do *not* relax between consecutive field modulation cycles, so χ' does not change sign, while χ'', which is due to nonperfect adiabaticity, changes sign.

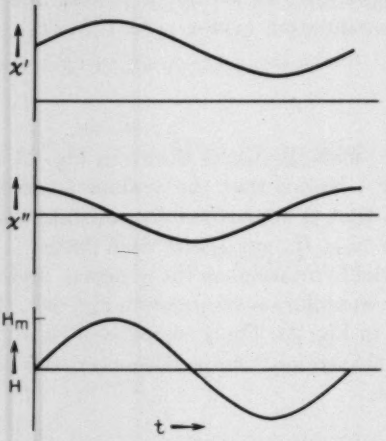


Fig. 28 — Scope traces of signal of case 5 (combination of the signals due to the individual spin packets, Fig. 27). Although χ' is large, its component at the field modulation frequency is small, and due to the derivative of the line shape.

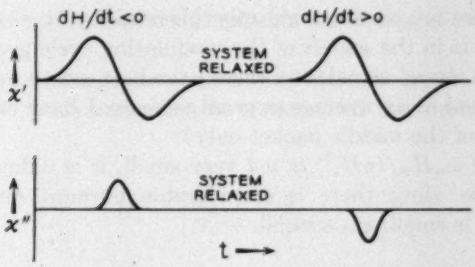


Fig. 29 — Expected recorder traces of signal of case 5.

packet is swept through, which varies with time due to the linear field sweep, dH_0/dt. It also depends upon the environment of each spin — and in an inhomogeneously broadened line, different packets will have different environments and different α's. Thus, only order of magnitude estimates of α, and of the amplitude, can be given.

In the limit when $\alpha N \ll 1$, the signal has been calculated by Portis.† Quoting his results, (with slightly changed notation):

$$\chi'(H) = \mp \tfrac{1}{2}\chi_0 H \int_0^\infty \frac{h(H_0 - H_0')}{\sqrt{(1+\delta^2)}} \frac{dH_0}{\gamma H_1} \approx \mp \frac{1}{2\sqrt{2\pi}} \chi_0 \frac{H}{\Delta H} \ln \frac{2\Delta H}{H_1}$$

for a gaussian line, with $H_1 \ll \Delta H$. Modulating the field, the component at the field modulation frequency is

$$\chi'\left(H_0' \pm \frac{\Delta H}{\sqrt{2}}\right) \approx \frac{1}{\sqrt{2\pi}} \chi_0 \frac{H_0}{\Delta H} \frac{H_m}{\Delta H} \ln \frac{2\Delta H}{H_1} \cos \omega_m t.$$

To include losses due to partial relaxation, multiply this value by

$$\frac{1 - e^{-\alpha N}}{\alpha N},$$

where α is the loss of magnetization per passage and

$$N = \frac{2}{\pi} \frac{H_m \omega_m}{dH_0/dt}.$$

The χ'' signal is calculated as in case 2, but, since the magnetization vector changes sign with dH/dt, χ'' changes sign too, and we have, neglecting burnout,

$$\chi'' = \chi_0 H \frac{H_m \omega_m \sin \omega_m t}{\alpha H_1^2} h(H - H_0').$$

† Ref. 5, p. 17.

Including burnout, we must multiply this result by $(1-\alpha)^{N/2} \approx e^{-\alpha N/2}$, since the packets in the *middle* of the modulating cycle give most of the contribution to the χ'' signal, and these have been swept over $N/2$ times only (i.e., instead of an average over-all packets, I have calculated here the amplitude of the middle packet only).

Note that, if $\omega_m H_m/(\gamma H_1^2)$ is *not* very small, it is difficult to get the 'nonburnt' case, since there is considerable burnout due to nonadiabaticity. If it is small, χ'' is small.

Experimental Data

See Fig. 30 (noting that there are *two* distinct lines on this trace).

Data: χ_f' (employing phase sensitive detection); $H_1 \approx 1/300$ gauss; $H_m = 1/20$ gauss; $\omega_m = 2\pi \times 100$ rad/second; $T = 1.2°$K. sweep rate: 8 gauss/second.

Conditions satisfied: H_1, $H_m \ll \Delta H$; $\gamma H_1 T_1 \gg 1$; $\gamma H_1^2 > \omega_m H_m$. Number of times each packet is swept:

$$\frac{2H_m}{dH/dt}\left(\frac{2\omega_m}{2\pi}\right) = 2.5$$

Loss of magnetization per passage: $\alpha \approx 0.1$ (roughly); $\gamma H_1^2 \approx 200$; $\omega_m H_m \approx 30$.

Interpretation: The recorder trace agrees quite well with the theoretically expected trace on Fig. 29.† Also, consecutive traces have someferent shapes and amplitudes. This will be discussed in Section 6.6.

6.6 *Case 6 — Rapid Adiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles, the Magnetization of the Spin Packets Being Partly Destroyed*

Physical Description

Each spin packet yields a rapid passage adiabatic signal. However, there is relatively more burnout than in case 5, and the burnout cannot

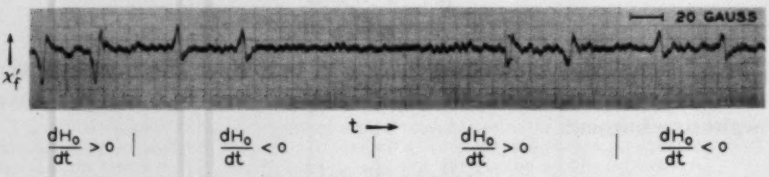


Fig. 30 — Experimental recorder trace of case 5 (compare with Fig. 29).

† Note that the system does *not* relax between consecutive sweeps, since T_1 is much larger than the time between sweeps.

be assumed just to reduce the amplitude of the signal without affecting its shape.

This case is an intermediate one and of little theoretical significance in itself. However, it is important to understand the processes occurring here. Let

$$H = \begin{cases} H_0 + H_m & t < 0 \\ H_0 + H_m \cos \omega_m t & t > 0 \end{cases}$$

(H_0 is at the center of the line.) Consider $\chi'(t)$ (see Fig. 31).

For $t \leqq 0$, $\chi'(t) = 0$, since we have the slow passage case, and are (say) at the center of the line.

For $0 < t < 1/\omega_m$ (say), we have a very fast rise of χ', since we have rapid passage through nonburnt packets. This is region I in Fig. 31.

For slightly larger t, we have a *constant* χ', independent of the sign of dH/dt. Since *both* dH/dt and S_z reverse their sign, χ' maintains its sign. This is region II.

Eventually, spin packets will burn out, either because of loss of magnetization *during* passage, or T_1 relaxation of "inverted" spins, between passages. The spins are partly burnt out in region III

Finally, a stationary state is reached, in which the burnout per passage is replaced by means of the T_1 relaxation, between passages (region IV).

Operation in regions I or II yields case 5. Operation in region IV yields case 7. Operation under conditions all regions contribute, but I, II, III contribute more than IV, yields case 8. Operation under conditions at which region III is important, yields this case, 6.

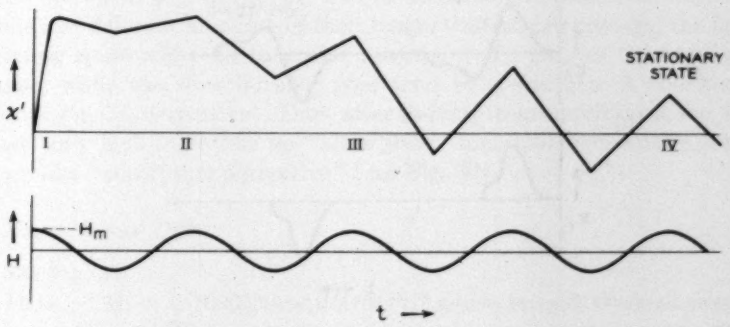


Fig. 31 — Transient χ' trace, for illustration of case 6: χ' is first large and constant; eventually, spin packets "burn out"; χ' decreases, but its component at the field modulation frequency may increase.

Conditions Under Which This Case Is Observed:

$\gamma H_1\sqrt{T_1T_2} \gg 1$; $\gamma H_1^2 \gg H_m\omega_m$; $\omega_m T_1 \gg 1$; $H_m \gg 1$; $N\alpha \approx 1 (N \equiv 2H_m/(dH_0/dt)2\omega_m/2\pi$; α: loss of magnetization per passage).

Traces

The χ' signal will be approximately as shown in Fig. 32(a). The χ'' signal will be as in case 5, but somewhat smaller, since, for packets in the middle of the modulation cycle, for which dH/dt is large and therefor χ'' (due to nonideal adiabaticity) should be appreciable, the magnetization is smaller.

Near the extrema of the modulation cycle we may have slow passage conditions and a corresponding χ'', which is, however, mostly at the second harmonic of the field modulation frequency. Thus, the output of

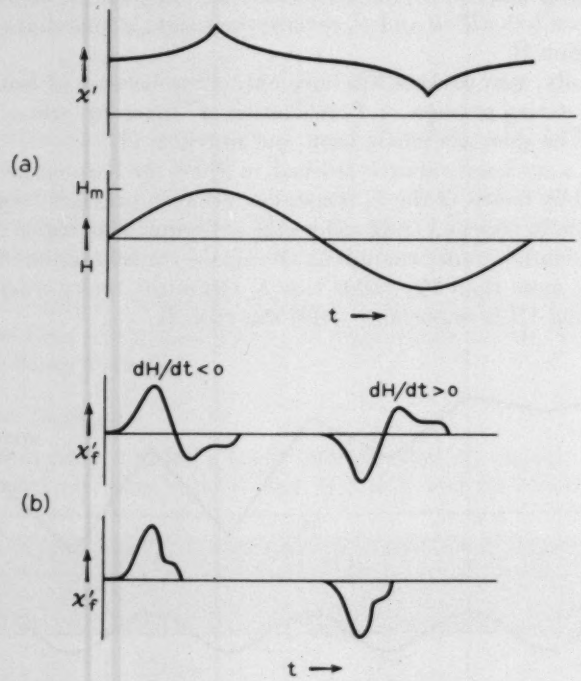


Fig. 32 — (a) Expected scope trace of case 6: spin packets just commence to "burn out"; χ' is large, but its component at the field modulation frequency is small. (b) Expected recorder traces of hybrids of case 5 with cases 6, 7, or 8.

the phase-sensitive detector is of "absorption" shape, 0° phase shift with field modulation for χ' and small for χ''.

A mixture of case 5 will distort the χ' signal and yield traces as shown in Fig. 32(b).

Amplitudes

Amplitudes are difficult to estimate, since the rate of burnout will be different for different spins and different parts of the field modulation cycle.

If the loss of magnetization is mainly due to T_1 relaxation between sweeps, the rate of burnout is about $1/T_1$, as discussed by Feher.†

The amplitude of the χ' signal will then be about that of case 7 (discussed below), but it may be larger or smaller.

If the burnout is mainly due to T_1 relaxation, the magnetization vectors at one extremum of the modulation cycle (dH/dt = maximum) will burn out fastest, and those at the other one will burn out slowly. Thus, a transient condition may develop in which the signal is *stronger* than in case 7.

If the spins at the extrema burn out fastest (since dH/dt is small there and "T_2" processes are therefore more effective) the signal will tend to be weaker than in case 7.

Therefore, attempting to give some general expression for the amplitude is not of much use here.

Comment

The mixture of cases 5 and 6 (and possibly 7 and 8) yields the peculiar trace shown in Fig. 32(b). If, due to different environments, different spins lose different amounts of their magnetization per passage, the fast-burning spins will tend to create case 6 (or 8), i.e., an "absorption" shape, while the slow burning ones tend to create case 5 (nonburnt absorption line derivative). Thus, after starting from equilibrium, the *first* trace may look more like an "absorption" line, and consecutive ones, more like "absorption derivative" (see Fig. 30).

Experimental Data

See Fig. 33.

Data: χ', $H_1 = 1/1000$ gauss; $dH/dt = \frac{1}{2}$ gauss/second. Over-all sweep: 1 gauss; $T = 4.5$°K.

† Ref. 12, p. 1242.

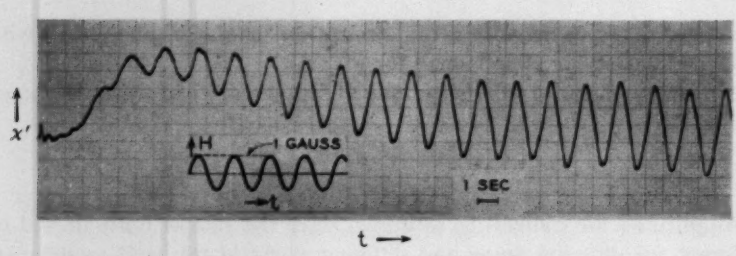


Fig. 33 — Transient χ' signal; this trace verifies experimentally the theoretically predicted trace of Fig. 31.

Conditions satisfied: H_1, $\delta H \ll \Delta H$; $\gamma H_1 T_1 \gg 1$; $\gamma H_1^2 \gg dH/dt$; "ω_m" $T_1 \gg 1$; ($\gamma H_1^2 \approx 20$; $\gamma H_1 T_1 \approx \frac{1}{2} \times 10^6$). Repetition rate: $\frac{1}{2}$ sec^{-1}; $T_1 \approx 5$ seconds.

Interpretation: This trace coincides with the theoretically predicted trace of Fig. 31.

6.7 *Case 7 — Rapid Adiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles — the Stationary Case*

Physical Description

Each spin packet yields an adiabatic rapid passage signal. However, the packets are in stationary state, which depends upon T_1 and the loss of magnetization per passage, but *not* upon dH_0/dt which is assumed to be so small as to let the stationary state be established. This case has essentially been discussed by Portis[5] [his Case III(b)] and Feher.[12]†

Conditions Under Which This Case Is Observed:

$$\gamma H_1 \sqrt{T_1 T_2} \gg 1; \qquad \gamma H_1^2 \geqq H_m \omega_m; \qquad \omega_m T_1 \gg 1; \qquad H_m \gg H_1.$$

In addition, the signal due to this case must be strong compared to that of cases 6 and 8, which compete with it. This requires: $dH_0/dt \ll H_m/T_1$ (see sections on amplitudes).

Traces

The spin packets, resonating at $H \approx H_0 - H_m$, are most of the time in a field *above* their resonant field. Therefore, their S_z will be positive for $H > H_0$, and thus negative for $H < H_0$.

† See Feher,[12] p. 1241–1242.

Thus, for $dH/dt < 0$, we get a positive χ', and for $dH/dt > 0$ a negative one. For packets resonating near $H \approx H_0 + H_m$ signs are reversed.

Packets resonating at $H \approx H_0$ are about half the time in a field above their resonant field, and half the time in a lower field. Therefore, their magnetization is small. We thus have the traces shown in Fig. 34.

Amplitudes

Let the fractional loss of magnetization during each passage be α. Then [as in Equation (33) of Feher[12]],

$$-S_N = [S_{N-1}e^{-t_1/T_1} - S_0(1 - e^{-t_1/T_1})](1 - \alpha),$$
$$-S_{N+1} = [S_N e^{-t_2/T_1} + S_0(1 - e^{-t_2/T_1})](1 - \alpha),$$

where S_N is the magnetization after the Nth sweep, and S_0 is the equilibrium magnetization.

In the stationary state:

$$\begin{aligned} S_{N-1} = S_{N+1} &= -[S_N e^{-t_2/T_1} + S_0(1 - e^{-t_2/T_1})](1 - \alpha) \\ &= e^{-t_2/T_1}[S_{N-1}e^{-t_1/T_1} - S_0(1 - e^{-t_1/T_1})](1 - \alpha)^2 \\ &\quad + S_0(1 - e^{-t_2/T_1})(1 - \alpha), \end{aligned}$$

$$S_{N-1} = \frac{e^{-t_2/T_1}(2 - \alpha) - e^{-(t_1+t_2)/T_1}(1 - \alpha) - 1}{1 - (1 - \alpha)^2 e^{-(t_1+t_2)/T_1}}(1 - \alpha)S_0,$$

$$S_N = \frac{e^{-t_1/T_1}(2 - \alpha) - e^{-(t_1+t_2)/T_1}(1 - \alpha) - 1}{1 - (1 - \alpha)^2 e^{-(t_1+t_2)/T_1}}(1 - \alpha)S_0.$$

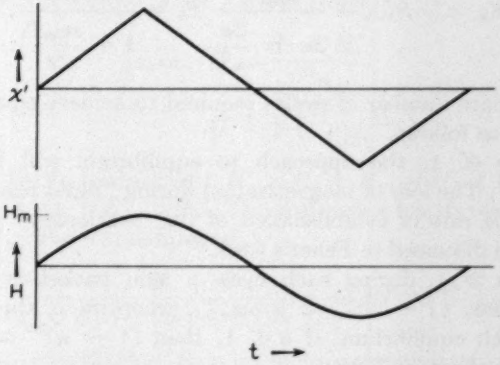


Fig. 34 — Scope trace of signal of case 7: a stationary state is established; the spin packets at the center of the field modulation cycle are burnt out, while those at the extrema yield an adiabatic rapid passage signal.

For $T_1 \gg t_1 + t_2$:

$$S_{N-1} \approx \frac{-(2-\alpha)\dfrac{t_2}{T_1} + (1-\alpha)\dfrac{t_1+t_2}{T_1}}{[1-(1-\alpha)^2] + (1-\alpha)^2\dfrac{t_1+t_2}{T_1}}(1-\alpha)S_0,$$

$$S_N \approx \frac{-(2-\alpha)\dfrac{t_1}{T_1} + (1-\alpha)\dfrac{t_1+t_2}{T_1}}{[1-(1-\alpha)^2] + (1-\alpha)^2\dfrac{t_1+t_2}{T_1}}(1-\alpha)S_0.$$

If, furthermore, $\alpha \ll 1$, then:

$$S_{N-1} = \frac{\dfrac{t_1-t_2}{T_1}}{2\alpha + \dfrac{t_1+t_2}{T_1}}S_0,$$

$$S_N = \frac{\dfrac{t_2-t_1}{T_1}}{2\alpha + \dfrac{t_1+t_2}{T_1}}S_0.$$

S_N, S_{N+1} are then still linear functions of $t_1 - t_2$, as is the case when $\alpha = 0$. For $t_2 = 0$, say,

$$S_N = -S_{N-1} = \frac{\dfrac{2\pi}{\omega_m T_1}}{2\alpha + \dfrac{2\pi}{\omega_m T_1}}S_0 = \frac{S_0}{1 + \dfrac{\alpha\omega_m T_1}{\pi}}.$$

The approximate number of cycles required to achieve equilibrium can be estimated as follows.

If $\alpha\omega_m T_1/\pi \ll 1$, the approach to equilibrium will be governed mainly by T_1. The loss of magnetization during "rapid passage" can be neglected. The rate of establishment of the "stationary" state will be about T_1^{-1}, as discussed in Feher's work.[12]

If $\alpha\omega_m T_1/\pi \gg 1$, during each cycle a spin packet is reduced by $(1-\alpha)^2$. Thus, $(1-\alpha)^{2m} \approx \pi/\alpha\omega_m T_1$, where m is the number of cycles to reach equilibrium. If $\alpha \ll 1$, then $(1-\alpha)^{2m} \approx e^{-2\alpha m}$, and $m \approx 1/(2a)\ \ln\ [\pi/(\alpha\omega_m T_1/\pi)]$.

Numerical examples (corresponding to conditions of the present experiments):

(a): $\alpha = 0.1$; $\omega_m = 2\pi \times 1000$ rad/second; $T_1 = 100$ seconds; $(\alpha\omega_m T_1)/\pi = 2000$; $m \approx 5(\log 2000) \approx 50$ cycles.

(b): $\alpha = 0.5$; $\omega_m = 2\pi \times 100$ rad/second; $T_1 = 5$ seconds; $(\alpha\omega_m T_1)/\pi = 500$; $4^m = 500$; $m \approx 5$ cycles.

[Packets near the center do not yield exactly zero signal, since when $S_z < 0$ relaxation is slightly faster than when $S_z > 0$. If we let $t_1 = t_2$,

$$S_N = S_{N+1} \approx -\frac{\alpha\dfrac{t_1 + t_2}{2T_1} + \frac{1}{4}\left(\dfrac{t_1 + t_2}{T_1}\right)^2}{2\alpha + \dfrac{t_1 + t_2}{T_1}} S_0,$$

and the signal due to these packets is of the second order. Spin packets at the center of the modulation cycle have been discussed in detail by Salpeter.[4]]

Using the result $S_N \approx \pi S_0/(\alpha\omega_m T_1)$, we have

$$\chi_f' \approx \frac{\chi_0 H_0}{\alpha\omega_m T_1} h(H_0' - H_0) \ln \frac{10H_m}{H_1}.$$

(χ'' is small).

This approximate result is derived as follows. The signal due to each packet is

$$\chi' = \frac{S_0}{2H_1} \frac{\pi}{\alpha\omega_m T_1},$$

where H' is the field at which the spin packet resonates and H_0 is the magnetic field at the instant at which the modulation field is zero. Superimposing the signals of the various packets,

$$\chi' \propto \int_{H_0 - H_m}^{H_0 + H_m} \frac{\dfrac{H_0 - H'}{H_m}}{\sqrt{1 + \left(\dfrac{H - H'}{H_1}\right)^2}} dH'$$

if the signal due to packets with $H' > H_0 + H_m$ or $H' < H_0 - H_m$ can be neglected (it can, if $H_m \gg H_1$). This expression has its maximum when $H = H_0 + H_m$. For it, the integral is

$$\frac{H_1}{H_m} \int_{H_0 - H_m}^{H_0 + H_m} \frac{H_0 - H'}{\sqrt{H_1^2 + (H_0 + H_m - H')^2}} dH' \approx H_1 \log \frac{4eH_m}{H_1}.$$

Now, since $S_0 = \chi_0 H h(H - H_0')$, we have, approximately,

$$\chi_f'(H) \approx \chi_0 H \frac{h(H - H_0')}{\alpha\omega_m T_1} \ln \frac{10H_m}{H_1}.$$

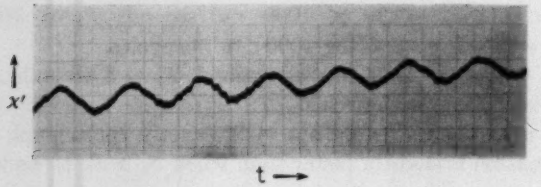


Fig. 35 — χ' of case 7 (no phase sensitive detector used.) Compare with Fig. 34.

Experimental Data

See Fig. 35.

Data: χ'; *no* phase-sensitive detection; $H_1 \approx 1/100$ gauss; "H_m" = 0.8 gauss; dH/dt = 0.8 gauss/second; T = 4.5°K (triangular sweep).

Interpretation: This trace corresponds to Fig. 34.

6.8 *Case 8 — Rapid Adiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles, the Magnetization of the Spin Packets Being Completely Destroyed*

Physical Description

In this case, the loss of magnetization per sweep is large, or T_1 is

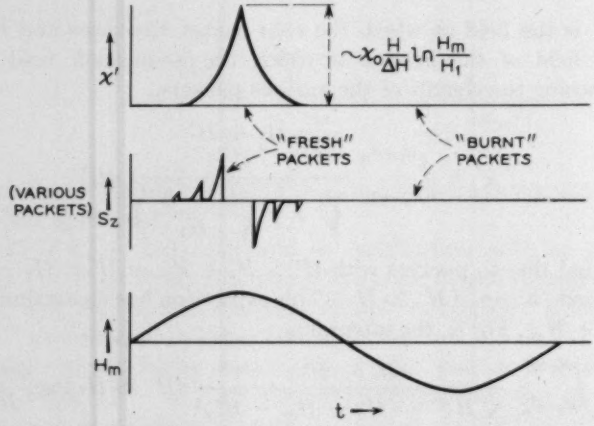


Fig. 36 — The contribution of the various spin packets to a "case 8" signal: spin packets near the extremum of the field modulation cycle have not been swept many times, and thus are not "burnt out"; packets further removed from the extremum are partly, or completely, burnt.

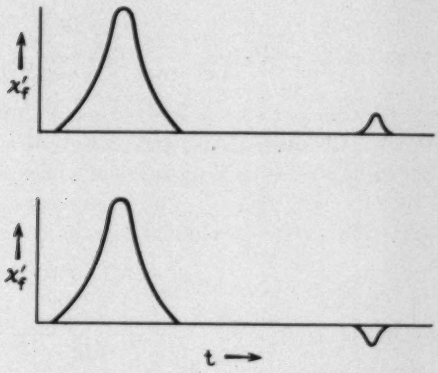


Fig. 37 — Expected recorder traces of case 8 signals: the line is burnt while sweeping through, consecutive sweeps yield weaker lines; their sign depends upon T_1, the time between consecutive sweeps, and the extent to which the line is "burnt out."

long, and the line burns out rather completely, and fast. Most signal is due to "fresh" packets covered by the advancing, modulated field.

Conditions Under Which This Case Is Observed:

$\gamma H_1 \sqrt{T_1 T_2} \gg 1$; $\gamma H_1^2 \geqq \omega_m H_m$; $\omega_m T_1 \gg 1$; $H_m \gg H_1$; $\alpha \omega_m \Delta H \gg dH_0/dt \gg H_m/T_1$; (the last condition ensures that the signal due to this case exceeds those of cases 5 and 7).

Traces

Since only the packets near the advancing extremum of the modulation cycle are not burnt, the χ' trace is as shown in Fig. 36. The output of the phase-sensitive detector is an "absorption" signal. When we sweep the line again after a time short compared with T_1, it is *saturated* and we get little response (Fig. 37). The χ'' signal again is small. It may be due to slow passage near the extremum, or nonadiabaticity; however, it is *not* at the second harmonic of the field modulation frequency.

Amplitudes

The χ' amplitude is

$$\chi_f' \approx \chi_0 H_0 h(H - H_0') \frac{dH_0}{dt} \frac{2\pi}{\alpha\omega_m} \ln\left(\frac{2\pi \frac{dH_0}{dt}}{\alpha\omega_m H_1}\right)$$

if the loss of magnetization per sweep is constant. (However, this is a very bad assumption.) This estimate is derived as follows:

If the burnout per cycle is α, after N passages of a packet its magnetization will be down by $(1 - \alpha)^N \approx e^{-\alpha N}$. Thus, the shape of the magnetization versus t curve, for a given modulation cycle, is approximately as shown in Fig. 38:

$$\chi_{\text{peak}}' \approx \chi_0 H h(H - H_0') \int_{-\infty}^{\infty} \frac{e^{-\left|\frac{\alpha\omega_m H'}{dH_0/dt\,\pi}\right|}}{\sqrt{1 + \left(\frac{H - H'}{H_1}\right)^2}} \frac{dH'}{H_1}$$

$$\approx 2\chi_0 H h(H - H_0') \ln\left(\frac{2\pi \frac{dH_0}{dt}}{\alpha\omega_m H_1}\right).$$

If the width of the distribution is small, the component of χ' at the field modulation frequency is about $(1/\pi)\chi_{\text{peak}}' \times$ (width of distribution). Or, about

$$\chi_f' \approx \frac{dH_0}{dt} \frac{2}{\alpha\omega_m H_m} (\chi_0 H) h(H - H_0') \ln\left(\frac{2\pi \frac{dH_0}{dt}}{\alpha\omega_m H_1}\right)$$

for $H_m \alpha\omega_m \gg dH_0/dt$.

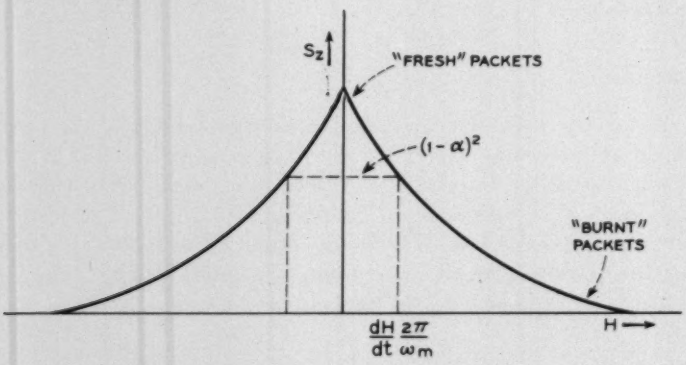


Fig. 38 — Magnetization in the range of one modulation cycle, case 8 (or 11): spin packets at the extremum of the field modulation cycle are not burnt out, while those removed from it are.

However, if the width becomes larger, there is no longer the linear dependence of χ' on dH_0/dt. In addition, α can by no means be considered constant. For $H_m\alpha\omega_m \approx dH_0/dt$, the width of the distribution will be of order H_m, and we may expect a signal of the order $\chi'_{\text{peak}}H_m$, or, very roughly, $\chi' \approx (\chi_0 H)h(H - H'_0) \ln H_m/H_1$.

Some experimental data concerning the dependence of χ' upon dH_0/dt and H_m are given in Section 6.12.

Experimental Data

See Fig. 39.

Data: $H_1 \approx 1/300$ gauss; $\omega_m = 2\pi \times 100$ rad/second; $H_m = 0.3$ gauss; $dH_0/dt = 3.2$ gauss/second; $T = 1.2°$K. Top trace is χ''; center and bottom traces are χ' with opposite values of dH_0/dt.

Conditions satisfied: $H_1, H_m \ll \Delta H$; $\gamma H_1 \sqrt{T_1 T_2} \gg 1$; $\gamma H_1^2 \geqq \omega_m H_m$; $\omega_m T_1 \gg 1$ ($\gamma H_1^2 \approx 200$; $\omega_m H_m \approx 200$).

Interpretation: These photographs correspond to the predicted traces of Fig. 36. The χ'' trace is weak and may be a spurious χ' signal (i.e., the bottom trace, weaker and inverted).

6.9 *Case 9 — Rapid Nonadiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles and Very Rapid Field Sweep*

Physical Description

In this case, each packet is traversed very rapidly, so that the passage

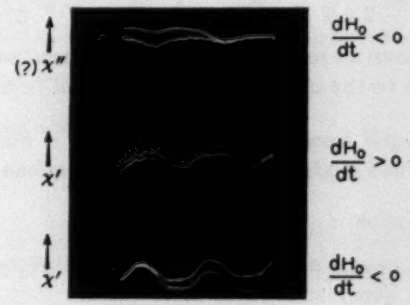


Fig. 39 — Experimental scope traces corresponding to case 8 (compare with Fig. 36): the χ'' trace is probably a spurious χ' signal; the signal here is weak, so the signal-to-noise ratio is not good.

is not adiabatic, and the burnout is small. The signal will be the superposition of "wiggles" treated in Section V.

If $T_2 \ll 1/\omega_m$,† the packets are initially aligned along the z-axis, and we may use the results of Section V straightforwardly. If T_2 is longer, there will be some "wiggles," but these will be rapidly averaged out by the limited bandwidth of the system. (For a more detailed treatment, see Appendix C.)

The χ'' signal is proportional to $h(H - H_0')$, so the component at the field modulation frequency is proportional to the derivative of the line shape and, similarly, that of χ' is roughly proportional to the second derivative of the line shape.

In the moderately nonadiabatic case, in the region for which $\gamma H_1^2 \approx dH/dt$, there is immediate burnout, and a "nonburnt" case is probably impossible.

Conditions Under Which This Case Is Observed:

$\gamma H_1^2 \lll \omega_m H_m$; $\gamma H_1 \sqrt{T_1 T_2} > 1$; $(\gamma H_1^2/(dH_0/dt)) \cdot (H_1/H_m) \lll 1$ (little burnout).

Traces

See Figs. 3(a) and 3(b).

Amplitudes

As in case 4,

$$\chi' \cong \frac{1}{2} \frac{H}{\Delta H} \chi_0 \frac{H_m}{\Delta H}; \qquad \chi'' \cong \frac{1}{2} \frac{H}{\Delta H} \chi_0 \frac{H_m}{\Delta H}.$$

This case requires an extremely low H_1 and fast sweep. Both these factors contribute to the degradation of the signal-to-noise ratio.

6.10 *Case 10 — Rapid Nonadiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles — the Stationary Case*

Physical Description

Each packet is passed almost suddenly. Equilibrium conditions have time to establish (very small dH_0/dt).

Note that if the passage is nonadiabatic, **S** is not inverted. All packages will have the same times to relax in the same direction.

† Here, T_2 is not T_{2*} (as at beginning of Section VI), but the zero of field T_2.

The fractional amount of magnetization cut off during a nonadiabatic passage is $\pi[\gamma H_1^2/(dH/dt)]$ (Section III). Thus, during those parts of the field modulation cycle in which the field varies fast, less will be cut off (see Fig. 40).

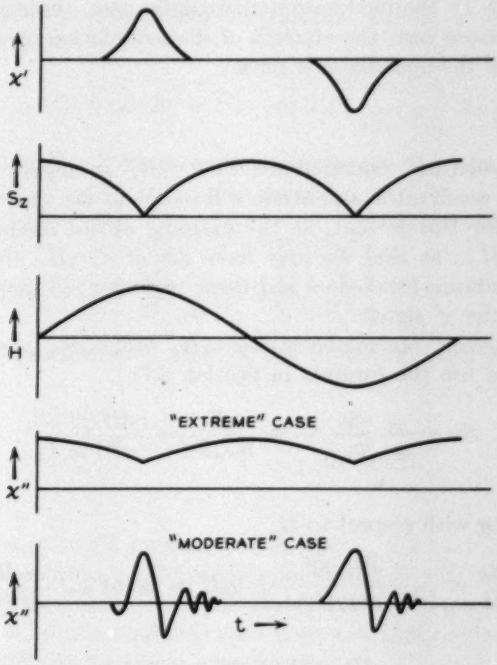


Fig. 40 — Expected scope traces of case 10: χ' depends critically upon *variations* of S_z ; these may be due to "burnout," or the derivative of the line shape; here, variations due to burnout are postulated.

The χ' signal has a sign depending on

$$\frac{dH}{dt}\frac{d|S|}{dH},$$

so that it will be as shown in Fig. 40, while χ'', which is proportional to $|S|$, will be mainly at the second harmonic frequency.

Conditions Under Which This Case Is Observed:

$\gamma H_1^2 \ll \omega_m H_m$; $\gamma H_1 \sqrt{T_1 T_2} \gg 1$; $\omega_m T_1(\gamma H_1^2/\omega_m H_m)$ not too large; $dH_0/dt \ll \omega_m H_m$; $H_m \ll \Delta H$.

(If this last condition is not fulfilled, we shall often have case 9, or a mixture of cases 9 and 10.)

Traces

See Fig. 40. In the moderate nonadiabatic case, the signal from the adiabatic portions near the extrema of the modulation cycle, may overwhelm that of the nonadiabatic ones.

Amplitudes

In the "moderate" nonadiabatic case (γH_1^2 is smaller than $\omega_m H_m$, but not very much), the amplitude will be about the same as in case 7. The reason for this is that, at the extrema of the modulation cycle, $dH/dt \ll \omega_m H_m$, so that we may have $dH/dt < \gamma H_1^2$ there, and the adiabatic condition fulfilled — and these regions yield most of the contribution to the χ' signal.

In the "extreme" nonadiabatic case ($\gamma H_1^2 \lll \omega_m H_m$), let $\alpha = \pi[\gamma H_1^2/(dH/dt)]$, and use the formula in Section 6.7:

$$S_N = \frac{S_0}{\frac{\alpha\omega_m T_1}{\pi} + 1} \approx \frac{\pi S_0}{\alpha\omega_m T_1} = \frac{|\, dH/dt \,|\, S_0}{\gamma H_1^2 \omega_m T_1}.$$

Differentiating with respect to H,

$$\frac{dS_N}{dH} = \frac{d^2H/dt^2}{(dH/dt)\gamma H_1^2 \omega_m T_1} \chi_0 H = \pm \cot \omega_m t \, \frac{\chi_0 H}{\gamma H_1^2 T_1}$$

since

$$H = H_0 + H_m \cos \omega_m t.$$

Note that this relation holds only for $dH/dt > \gamma H_1^2$, or

$$\omega_m H_m \,|\, \sin \omega_m t \,| > \gamma H_1^2, \quad |\, \sin \omega_m t \,| > \frac{\gamma H_1^2}{\omega_m H_m}.$$

From the result of Section V,

$$\chi = \tfrac{1}{2} H \chi_0 \frac{d\text{“}h\text{”}}{dH} \Delta H,$$

where "h" may be any "line shape" — here it is the "shape" due to the burnout of spin packets — and

$$\frac{d\text{“}h\text{”}}{dH} = h(H - H_0') \cot \omega_m t \, \frac{1}{\gamma H_1^2 T},$$

where $h(H - H_0')$ (without quotation marks) is the true line shape. Thus,

$$\chi' = \tfrac{1}{2}H\chi_0 h(H - H_0') \cot \omega_m t \left(\frac{1}{\gamma H_1^2 T_1}\right) \times \text{ width of distribution.}$$

The width of the magnetization distribution will be of order H_m (see Fig. 40), so that

$$\chi' \approx \tfrac{1}{2}H\chi_0 h(H - H_0') \cot \omega_m t \left(\frac{1}{\gamma H_1^2 T_1}\right) H_m ,$$

and the component at the modulation frequency is roughly

$$\chi_f' = \tfrac{1}{2}H\chi_0 h(H - H_0') \frac{H_m}{H_1} \frac{1}{\gamma H_1 T_1} \int_{\gamma H_1^2/(\omega_m H_m)}^{\pi/2} \cot \omega_m t \cos \omega_m t \, d(\omega_m t)$$

$$= \tfrac{1}{2}H\chi_0 h(H - H_0') \frac{H_m}{H_1} \frac{1}{\gamma H_1 T_1} \ln\left(\frac{2\omega_m H_m}{e\gamma H_1^2}\right).$$

Here, χ'' is approximately that of the nonburnt case, multiplied by $\pi/(\alpha\omega_m T_1)$:

$$\chi'' \approx \chi_0 H h(H - H_0') \frac{\pi}{\gamma\omega_m T_1} \frac{H_m}{\Delta H}.$$

Comment

If γH_1^2 is very small compared with $\omega_m H_m$, we have little burnout even near the extrema of the modulation cycle. If γH_1^2 is of the order $\omega_m H_m$ (or slightly smaller), the passage near the extrema of the modulation cycle is adiabatic, and this case does not apply either. Therefore, this case is difficult to create experimentally.

6.11 *Case 11 — Rapid Nonadiabatic Passage with a Short Time Between Consecutive Field Modulation Cycles, the Magnetization of the Spin Packets Being Completely Destroyed*

Physical Description

In this case, the adiabatic condition $\gamma H_1^2 > dH/dt$ is violated. In addition, the magnetization is destroyed during the passage, and most of the contribution to the signal comes from spin packets that have not been swept over many times. The general considerations concerning the "stale" case 10, apply here too.

Again, we must distinguish between the "moderately" nonadiabatic case, in which the passage near the extrema of the modulating field is

adiabatic, and the extreme nonadiabatic case in which γH_1 is *very* small compared with $\omega_m H_m$.

The moderate nonadiabatic case is essentially the same as case 8, the adiabatic fresh case. This case applies, if the width of the nonburnt region (Fig. 38) is limited by the adiabatic condition $dH/dt = \gamma H_1^2$, or if it is narrower. Since the burnout α is a very indefinite quantity, a theoretical estimate is very unreliable. However, the width of the burnt region can *very* easily be determined from the oscilloscope trace.

The extreme nonadiabatic case is much more difficult to treat. It should be noted that, near the extrema of the modulation cycle, dH/dt is extremely nonuniform, and it is not always clear whether solutions obtained assuming *constant* dH/dt apply here. Some of the considerations are given in Appendix C.

Conditions Under Which This Case Is Observed:

$\gamma H_1^2 < \omega_m H_m$; $\gamma H_1 \sqrt{T_1 T_2} \gg 1$; $\omega_m T_1 \gg 1$; $\gamma H_1^2/\omega_m H_m$ not too small compared with unity (otherwise we have case 9) $(dH_0/dt)/\omega_m H_m$ not too small compared with unity (otherwise we have case 9 or 10).

Traces

For the extreme nonadiabatic case, see Appendix C. For the moderately nonadiabatic case, the traces are exactly the same as for case 8 (Figs. 36 and 37).

Amplitudes

For the moderate case, the amplitudes are the same as for case 8. The signal of the extreme case is difficult to estimate. For some considerations, see Appendix C.

Experimental Data

(*a*) *Fig. 41(a).*

Data: χ' Signal; $H_1 \approx 1/100$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $H_m = 1$ gauss. Sweep rate: 1 gauss/second; $T = 1.2°$K. Top trace: light off, opposite signs of dH_0/dt; center and bottom traces: light shining on the sample, relaxing it (see Feher[12]).

Conditions satisfied: $\gamma H_1 T_1 \gg 1$; $\omega_m T_1 \gg 1$; $\omega_m H_m > \gamma H_1^2$. Number of times each packet is swept through: about 4000 ($\gamma H_1^2 \approx 2000$; $\omega_m H_m \approx 6000$).

Interpretation: This is a "moderately" nonadiabatic case; i.e., $\gamma H_1^2 <$

$\omega_m H_m$ but not $\gamma H_1^2 \lll \omega_m H_m$. At the extrema of the modulation cycle, the adiabatic condition is satisfied, and thus the trace of this case is indistinguishable from case 8 (see Fig. 39).

(*b*) *Fig. 41(b)*.

Data: $H_1 \approx 1/1000$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $H_m = 1/10$

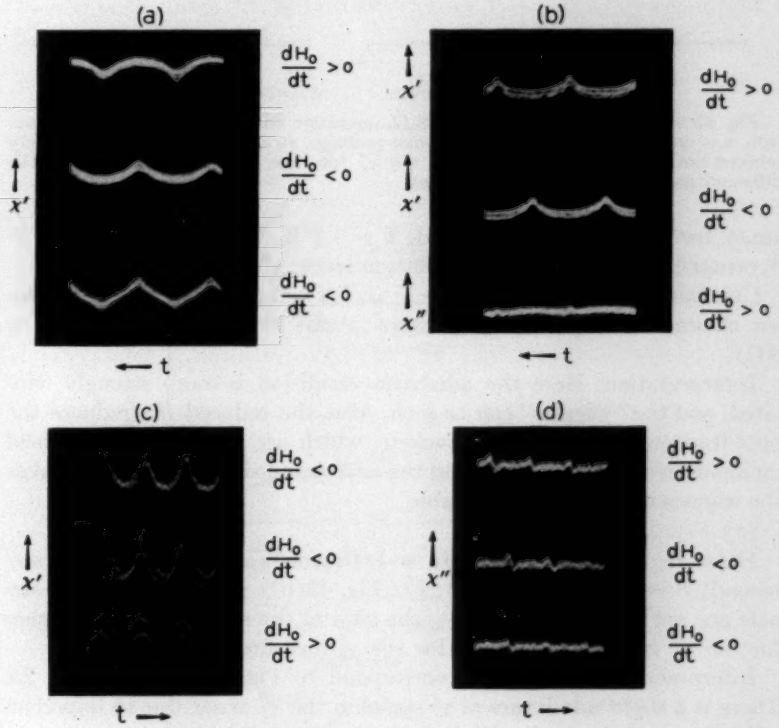


Fig. 41 — (a) Top and center traces: experimental scope χ' trace corresponding to a "moderate" case 11; ($\omega_m H_m > \gamma H_1^2$, but, only the spin packets near the extrema of the modulation cycle are not "burnt," and for these $dH/dt < \gamma H_1^2$); this case is very similar to case 8 — compare with Figs. 36 and 39. Bottom trace: corresponds (essentially) to case 7 (see Fig. 34) indicating that light reduces T_1, but it is still long compared with a field modulation cycle. (b) Experimental scope χ' and χ'' traces of an 'extreme' case 11 ($\gamma H_1^2 \ll \omega_m H_m$); the 'wiggles' are an indication of the nonadiabatic conditions near the extremum of the modulation cycle, where dH/dt is small. (c) χ' scope trace corresponding to case 11, the conditions being very similar to those of (a); however, a lower power reduced losses due to forbidden lines ($\gamma H_1^2/\omega_m H_m$ is the same). Data: $H_1 \approx 1/300$ gauss; $H_m = 1/10$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 1.2°$K. Top trace: light shining on the sample; center and bottom traces: no light shining on the sample; opposite signs of dH_0/dt. (d) χ'' scope trace corresponding to case 11. Data: $H_1 \approx 1/300$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 1.2°$K. Top and center traces: light shining on the sample opposite senses of dH_0/dt; bottom trace: no light shining on sample.

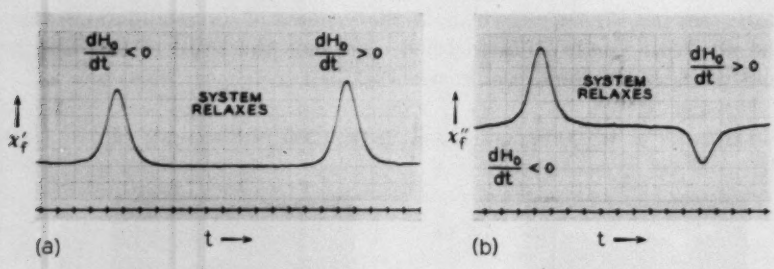


Fig. 42 — (a) χ_f' for case 11, with P.S.D. (compare with Fig. 37); the magnetization was completely burnt out after one passage, so the sample was completely relaxed between consecutive sweeps. (b) χ_f'' for case 11, with P.S.D. Note the different parity of the χ' and χ'' traces.

gauss. Sweep rate: 8 gauss/second; $T = 1.2°$K. Top trace: χ', $dH_0/dt > 0$; center trace: χ', $dH_0/dt < 0$; bottom trace: χ''.

Conditions satisfied: $\gamma H_1 T_1 \gg 1$; $\omega_m T_1 \gg 1$; $\gamma H_1^2 \ll \omega_m H_m$. Number of times each spin packet is swept: about 50 ($\gamma H_1^2 \approx 20$; $\omega_m H_m \approx 600$)

Interpretation: Here the adiabatic condition is more strongly violated, and the "wiggles" can be seen. Also, the reduced H_m reduces the beat frequency between spin packets, which are a fraction of the field modulation cycle out of tune and the microwave signal. This, too, makes the wiggles more easily observable.

(*c*) *Fig. 42.*

Data: $H_1 \approx 1/1000$ gauss; $H_m = 1/10$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 1.2°$K. Fig. 42(a): χ_f'; Fig. 42(b): χ_f''. The χ_f' and χ_f'' signals are *not* to the same scale — the gain of the system is 10 db higher for the χ_f'' measurement than for the χ_f' measurement.

Interpretation: These data correspond to Figs. 3(c), 3(d) and 25. There is a slight admixture of χ_f' signal in the χ_f'' trace, due to imperfect adjustment of the bridge. We have a "moderate" nonadiabatic case.

6.12 *Conditions to Maximize Signal*

An experiment was carried out to determine the dependence of the amplitude of the χ_f' signal upon H_1, H_m and dH_0/dt, for cases 9, 11 and neighboring ones.

The 11 cases treated in this section are idealizations, and usually we shall have a case intermediate between them. Thus, quantitative agreement with the formulas given in this section cannot be expected. However, there seems to be qualitative agreement with the theory. The results of the experiments are shown in Fig. 43.

From the dependence of the observed signal strength upon H_1 and H_m it is clearly seen that in order to obtain a strong signal, the adiabatic condition $\gamma H_1^2 > dH/dt$ should be violated, but not by too much (say, roughly $3 < (dH/dt)/\gamma H_1^2 < 10$).

If it is not violated, we do not have much burnout, and thus have cases 5 or 6, which yield quite low signals. If the adiabatic condition is

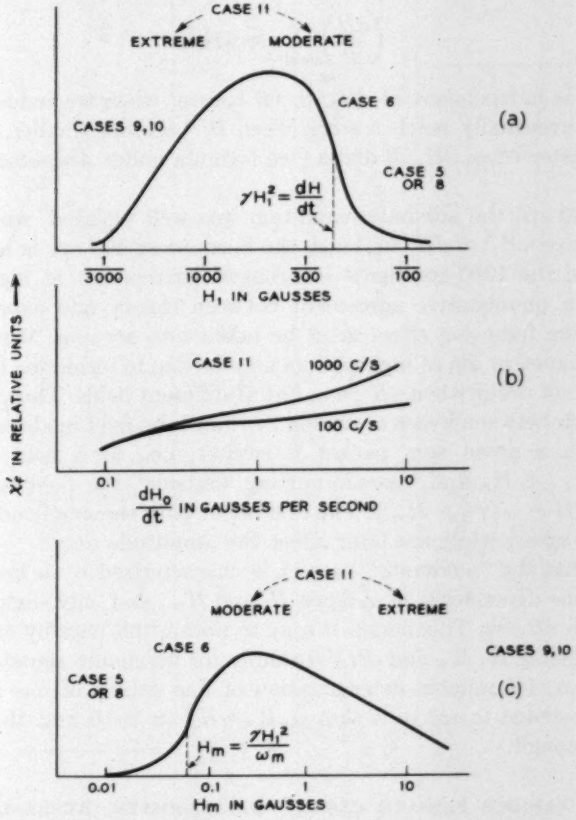


Fig. 43 — Dependence of amplitude of χ_f' signal (as detected by phase sensitive detector) upon H_1, H_m, dH_0/dt, for case 11: (a) Dependence upon H_1. Data: $H_m = 1/10$ gauss; $\omega_m = 2\pi \times 100$ rad/second; $dH_0/dt = 2$ gauss/second; $T = 1.2°K$; concentration, 3.2×10^{15} phosphorus atoms/cc. (b) Dependence upon dH_0/dt. Data: $H_1 = 1/300$ gauss; $H_m = 1/10$ gauss; $T = 1.2°K$; concentration, 3.2×10^{15} phosphorus atoms/cc. (c) Dependence upon H_m. Data: $H_1 \approx 1/300$ gauss; $\omega_m = 2\pi \times 100$ rad/second; $dH_0/dt = \frac{1}{2}$ gauss/second; $T = 1.2°K$; concentration, 3.2×10^{15} phosphorus atoms/cc.

violated too much, we again do not have much burnout, since we have an almost sudden transition (see Fig. 11). We then approach case 9 and case 10, which again yield weak signals.

The dependence of the signal strength upon dH_0/dt is weak. This is due to the fact that, when the burnout is due mainly to the breakdown of the adiabatic condition, the width H' of the unburnt region is given by

$$\left(\frac{dH}{dt}\right)_{H=H'} \approx \gamma H_1^2,$$

and this is independent of dH_0/dt. Of course, when we reduce dH_0/dt, we shall eventually reach a stage when H' becomes smaller, and then χ_f' falls faster when dH_0/dt drops (see formula under *Amplitudes* in Section 6.8).

At 1000 cps the adiabatic condition was well violated, while at 100 cps we have $\gamma H_1^2 \approx dH/dt$. Thus, the burnout at 100 cps is not so efficient, and the 1000-cps signal is stronger, particularly at high dH_0/dt.

When a quantitative agreement between theory and experiment is desired, the following effect must be taken into account. Much of the loss of magnetization at high powers may be due to forbidden lines. This loss does not occur when $\gamma H = \omega$, but at different fields. Thus, one must distinguish between losses occurring "within" the field modulation cycle (at which a given spin packet is swept), i.e., at a field satisfying $|H - \omega/\gamma| < H_m$ and losses occurring "outside" the field modulation cycle, ($|H - \omega/\gamma| > H_m$). The first affect both the amplitude and the "passage case", while the later affect the amplitude only.

Note that the "moderate" case 11 is characterized by a broad maximum in the dependence of χ_f' upon H_1 and H_m, and only slight dependence upon dH_0/dt. This makes it easy to obtain this case by experimentally adjusting H, H_m and dH_0/dt simply for maximum signal.

However, for reliable determination of fine details of line shapes, it is recommended to see to it that $\omega_m H_m/\gamma H_1^2 \approx 3$–10 and that dH_0/dt is large enough.

VII. THE VARIOUS PASSAGE CASES — DIAGRAMMATIC REPRESENTATIONS AND ADDITIONAL EXPERIMENTAL POINTS

In this section, we attempt to present some hints that may be of aid in the experimental identification of "passage" cases. The section contains a number of diagrams, suggestions and experimental data.

Table II indicates the logical relationship of the various cases, and scope and recorder traces to be expected.

TABLE II — LOGICAL RELATIONSHIP OF THE VARIOUS PASSAGE CASES

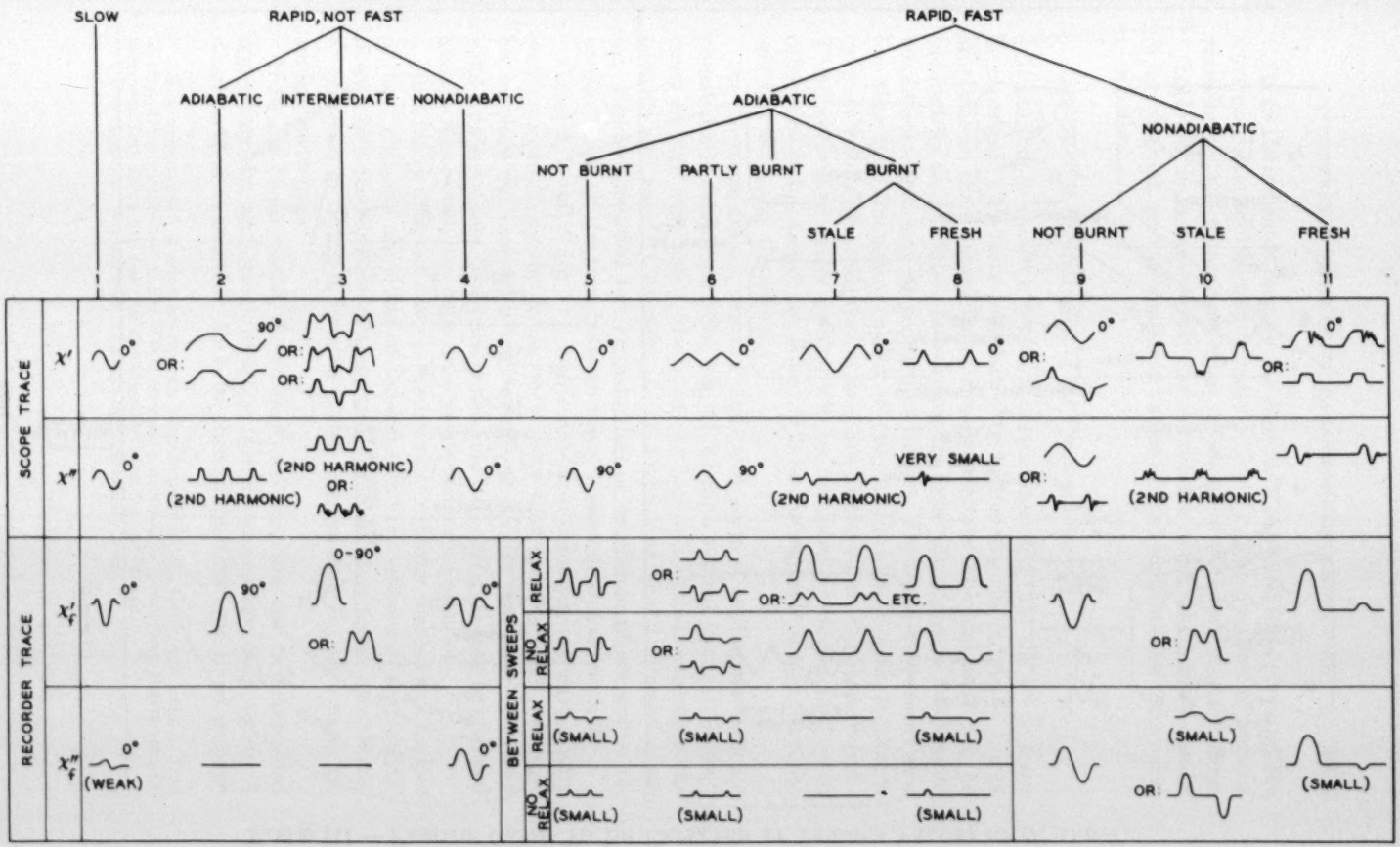


TABLE III — PASSAGE CASES TO BE EXPECTED AT VARIOUS VALUES OF H_1 AND T_1

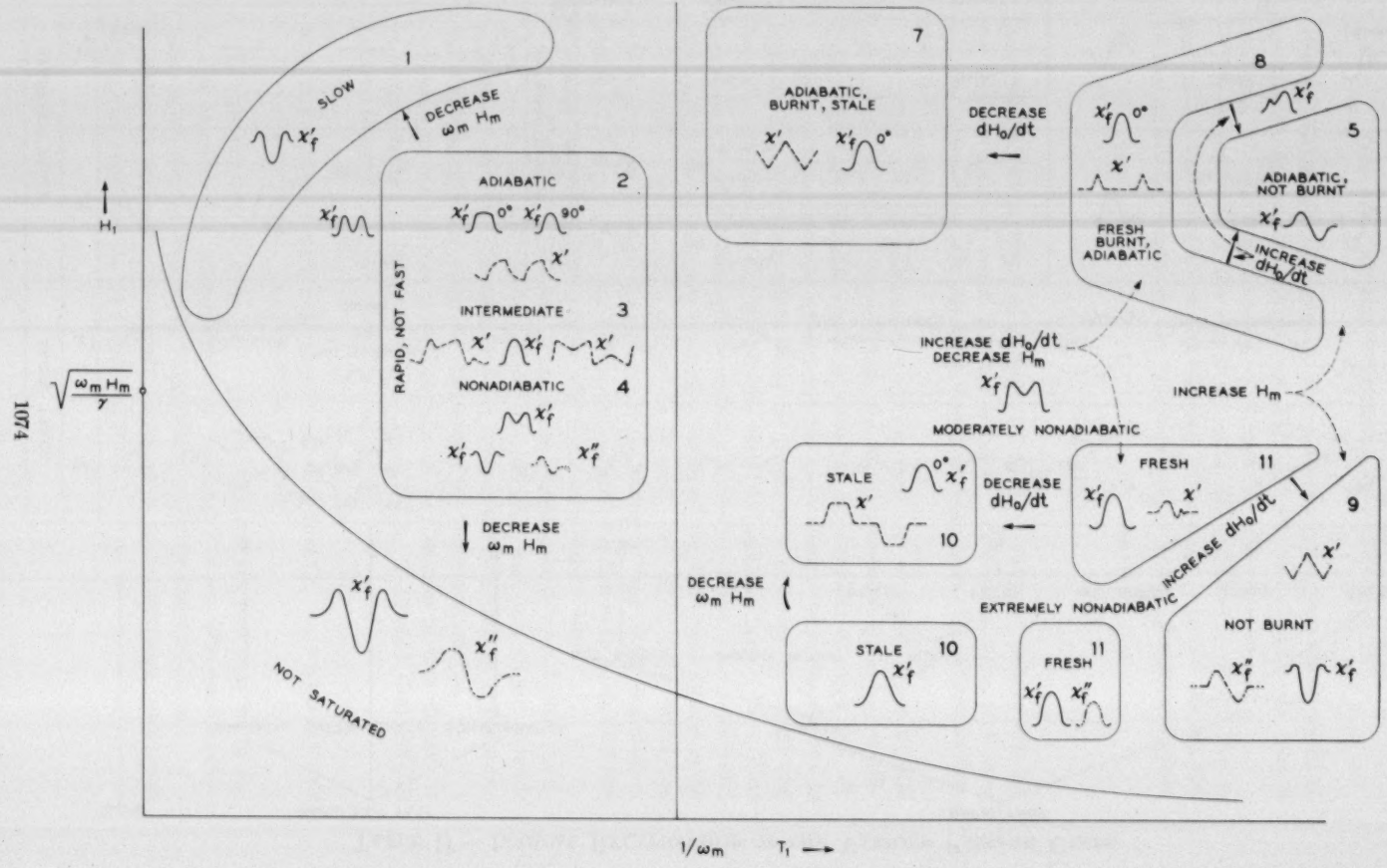


Table III indicates roughly the case to be expected for a given microwave magnetic field H_1 and relaxation time T_1, and the effect of various variations of the parameters upon the shape of the experimental traces.

Table IV indicates roughly the conditions under which each case is to be expected, and approximate values of the amplitudes of the χ_f' and χ_f'' signals.

Experimental Identification of the Passage Case

If nothing is known *a priori*, it is recommended that the scope trace be observed. This trace is obtained by connecting the output of the IF amplifier to the Y amplifier of the oscilloscope. Thermal noise can be reduced by an RC filter. Low-frequency noise often proves much more troublesome and may be reduced as follows: The scope is operated at a *low* time base frequency (about 10 cps) and a high X gain, so that a cycle or two of the modulating field are displayed on the scope screen. Then, successive cycles (which will be displaced vertically from each other, due to low-frequency noise), will be distinguishable and the noise will not wash out the signal.†

On the scope trace, "wiggles" and other sorts of oscillations indicate nonadiabatic conditions.‡ See Fig. 41(b) for a typical picture. "Fresh" cases are also very easily identified (by lack of symmetry between the positive and negative extrema of the modulating field.) "Fresh" cases can also be identified by the reduction of amplitude of successive recorder traces (if the time between them is short compared with T_1.)

Cases 2, 7 and 8 (or 11), which are most common, are particularly easy to identify by the scope trace. Note that the sweep dH_0/dt must be *on*, since the case occurring depends on dH_0/dt, and altering it may yield a different case. The parameters (H_m, H_1, dH_0/dt) may be changed and the effect upon the scope trace yields further information.

Since many different cases possess similar recorder traces (as a matter of fact, χ' traces are either of "absorption" or "dispersion derivative" shapes, or a combination of these two, commonly), extreme care should be exercised in trying to identify a case (or measure T_1, say) using recorder traces *only*.

If H_1 can be estimated, the passage case occurring can be determined approximately, and resort to scope traces can perhaps be avoided. A *direct* method to estimate H_1 is to photograph adiabatic rapid passage

† The appearance of such traces can be seen in Figs. 39, 41(c), and others. Due to the phosphorescense of the CRT screen, a number of preceding traces can be seen too, but these are easy to distinguish from the last trace.

‡ For these to be observed, T_2 must be sufficiently long.

TABLE IV — CONDITIONS UNDER WHICH THE VARIOUS PASSAGE CASES APPLY AND APPROXIMATE SIGNAL AMPLITUDES

Conditions			Case Number	x_f'	x_f''
$\omega_m T_1 \ll 1$	$\gamma H_1^2 > \omega_m H_m$	$T_1 < \dfrac{H_1}{\omega_m \Delta H}$	1	$\chi_0 \dfrac{H_0}{\Delta H} \dfrac{H_m}{\Delta H}$	$\chi_0 \dfrac{H_0}{\Delta H} \dfrac{1}{H_1\sqrt{T_1 T_2}} \dfrac{H_m}{\Delta H}$
		$T_1 > \dfrac{H_1}{\omega_m H_m}$	2	$\chi_0 \dfrac{H_0}{\Delta H} \ln \dfrac{H_m \omega_m T_1}{H_1}$	$\chi_0 \dfrac{H_0}{\Delta H} \dfrac{H_m}{\Delta H} \dfrac{H_m}{H_1} \dfrac{\omega_m}{\gamma H_1}$
	$\gamma H_1^2 < \omega_m H_m$		3	$\chi_0 \dfrac{H_0}{\Delta H} \ln \dfrac{H_m \omega_m T_1}{H_1}$	
	$\gamma H_1^2 \ll \omega_m H_m$		4	$\dfrac{1}{2} \dfrac{H_0}{\Delta H} \chi_0 \dfrac{H_m}{\Delta H}$	$\dfrac{1}{2} \dfrac{H_0}{\Delta H} \chi_0 \dfrac{H_m}{\Delta H}$
	$\gamma H_1^2 > \omega_m H_m$	$\dfrac{\omega_m H_m}{dH_0/dt} \alpha \ll 1$	5	$\chi_0 \dfrac{H_0}{\Delta H} \dfrac{H_m}{\Delta H} \ln \dfrac{2\Delta H}{H_1}$	$\chi_0 \dfrac{H_0}{\Delta H} \dfrac{H_m \omega_m}{\gamma H_1^2}$
		$\dfrac{\omega_m H_m}{dH_0/dt} \alpha \approx 1$	6		small
		$\dfrac{dH_0}{dt} \ll \dfrac{H_m}{T_1}$	7	$\chi_0 \dfrac{H_0}{\Delta H} \dfrac{\ln \dfrac{H_m}{H_1}}{\alpha \omega_m T_1}$	small

		$\frac{H_m}{T_1} \ll \frac{dH_0}{dt} \ll \alpha\omega_m\Delta H$	8	$(\chi_0 H_0/\Delta H)\ \ln H_m/H_1$ or $\chi_0 \frac{H_0}{H} \frac{dH_0/dt}{\alpha\omega_m} \ln \frac{dH_0/dt}{\alpha\omega_m H_1}$	small
$\omega_m T_1 \gg 1$		$\frac{\gamma H_1^2}{dH_0/dt} \frac{H_1}{H_m} \lll 1$	9	$\frac{1}{2} \frac{H_0}{\Delta H} \chi_0 \frac{H_m}{\Delta H}$	$\frac{1}{2} \frac{H_0}{\Delta H} \chi_0 \frac{H_m}{\Delta H}$
	$\gamma H_1^2 < \omega_m H_m$	$\frac{dH_0}{dt} \ll \omega_m H_m$ $\omega_m T_1 \frac{H_1^2}{\omega_m H_m} \leqq 1$	10	$\chi_0 H_0 \frac{H_m}{\Delta H} \frac{\ln \omega_m H_m/\gamma H_1^2}{\gamma H_1^2 T_1}$	$\chi_0 \frac{H_0}{\Delta H} \frac{H_m}{\Delta H} \frac{\pi}{\alpha\omega_m T_1}$
		$\frac{dH_0}{dt} \approx \omega_m H_m$ $\omega_m T_1 \frac{H_1^2}{\omega_m H_m} > 1$	11	$\chi_0 \frac{H_0}{\Delta H} \ln \frac{H_m}{H_1}$ or $\chi_0 \frac{H_0}{\Delta H} \frac{dH_0/dt}{\alpha\omega_m} \ln \frac{dH_0/dt}{\alpha\omega_m H_1}$	small

χ' and χ'' transients (if T_1 is large). The ratio of χ'' to χ' peaks is

$$\frac{\frac{dH}{dt}}{\gamma H_1^2} \frac{1}{\ln \frac{2\Delta H}{H_1}}$$

if $\Delta H < (dH/dt)T_1$; otherwise, the ratio is

$$\frac{\frac{dH}{dt}}{\gamma H_1^2} \frac{1}{\ln \frac{2\frac{dH}{dt}T_1}{H_1}}$$

(see Section 6.2, and Appendix D.) From this relation, H_1 can be determined (H_1 occurs under the logarithm too, but there its effect is not large, and a crude approximation will do.)

These photographs can be obtained by letting $H_m > \Delta H$, (say $H_m = 5\Delta H$), switching H_1 off, letting the system relax, switching H_1 on and letting the voltage that switches H_1 on, trigger the scope (switched to "one-shot" operation).

Another way is to set H_m to an extreme position ($+$ or $-$) off the line, letting the line relax, and then switching the modulating field "on."

If T_1 is short (but not too short), it is possible to avoid the use of transients and use steady-state operation. Then, amplitudes can be read from the scope directly, without use of a camera. Of course, H_1 must be in the right range ($\gamma H_1^2 > dH/dt$, but not $\gamma H_1^2 \gg dH/dt$). Since $H_m \gg \Delta H$, dH/dt equals $\omega_m H_m$; thus, γH_1^2 can be determined quite accurately.

If signals at twice the modulation frequency can be detected, χ'' signals due to slow passage near the extrema of the modulation cycle, or due to nonadiabaticity in the middle of the modulation cycle, can be observed. Also, "fresh" cases yield strong second harmonic signals, while "stale" and "nonburnt" cases yield weak second harmonic signals, as illustrated in Fig. 44.

It has been observed experimentally that, if the conditions are adjusted for maximum amplitude, the passage conditions often correspond to case 11.

VIII. CONCLUSION

Abragam[16] states that "If the line is inhomogeneously broadened with a width ΔH, and the modulation and the scanning are defined by

a law $H = H_0(t) + H_m \cos \Omega t$, no less than seven parameters having the dimensions of frequency may be introduced to discuss the shape and the size of the signals observed; namely, $1/T_1$, $1/T_2$, $\gamma \Delta H$, γH_1, γH_m, Ω, $(1/H_1)(dH_0/dt)$. The shape of the signals depends considerably on the relative values of these parameters and may become very complicated. They can, in practice, be analyzed mathematically using Bloch equations, and a considerable literature exists on this subject. However, there is nothing fundamental in the complications of these shapes, and the rather negative conclusion to be drawn from their analysis is the following: whenever the methods of observation are such that a complicated mathematical treatment is required to establish a relationship between the data (signals) and the physical nature of the system studied, the methods of observation are inadequate and must be changed, whenever possible."

The aim of this work is *not* to suggest using the complicated expressions given here to determine T_1, T_2, etc. The expressions are at best of qualitative value. This work attempts rather

i. to determine whether the seven parameters are sufficient to determine the signals (i.e., to determine whether the physical system studied can be described to a sufficient accuracy by the Bloch equations), and

ii. to determine whether line shapes, relative amplitudes, etc., can be determined in a reliable way under certain (complicated) relationships

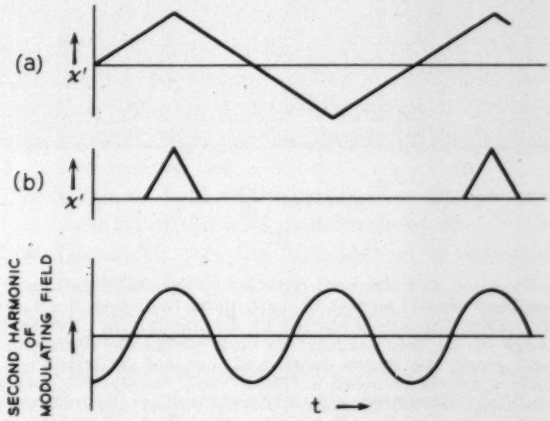


Fig. 44 — Demonstration of the strong harmonic content of "fresh" signals: (a) χ' of case 7 (stale); (b) χ' of case 8 (fresh). Note that the "stale" signal possesses only odd harmonics of the magnetic field modulation frequency, while the "fresh" signal possesses a strong second harmonic component.

between the variables (under conditions that insure a good signal-to-noise ratio) and, if they can, to determine the right conditions for this purpose.

Concerning i, it was found that forbidden lines had a drastic effect upon the signals observed here. With other samples, spin diffusion processes might become very important. Therefore, inhomogeneously broad-

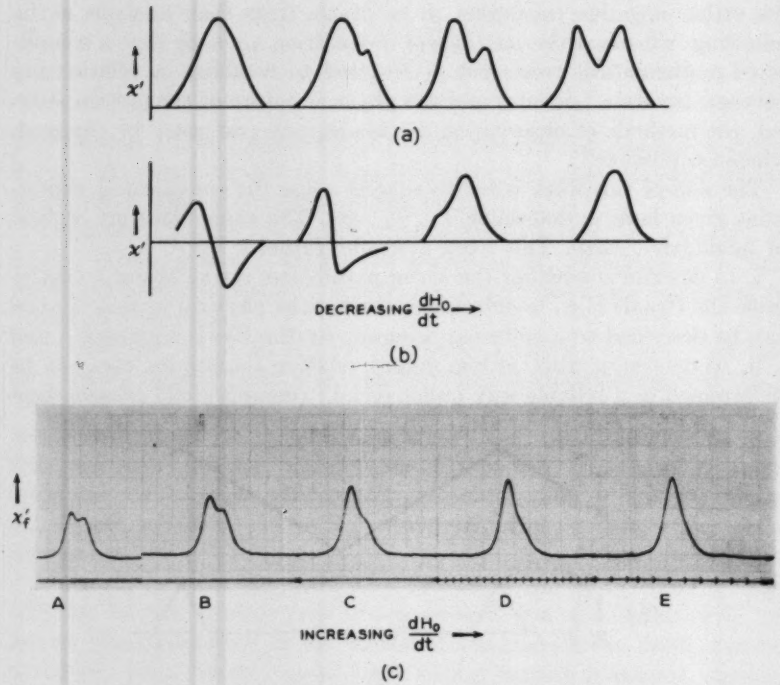


Fig. 45 — Expected and observed recorder χ_f' signals of various hybrid cases: (a) Transition from case 11 to case 10 (or 9) upon reduction of dH_0/dt. Note that, in the "fresh" cases, the line is shifted by an amount H_m, due to the fact that the *leading* edge of the field modulation cycle yields the signal; in the "stale" and "unburnt" cases, the line is broadened, but not shifted. Thus, for $H_m/\Delta H$ not very small, a hybrid between a "fresh" and "nonfresh" case will yield an unsymmetrical line; increasing $\omega_m H_m/\gamma H_1^2$ will *reduce* the burnout and tend to produce case 9; increasing $H_m/\Delta H$ has the same effect. Therefore, increasing H_m has a very large effect in producing "splitting" of a line. (b) Transition from case 5 to case 8 (or 7) upon decrease of dH_0/dt. (c) Experimental change of shape of χ' trace upon increase of dH_0/dt. Data: $H_1 \approx 1/300$ gauss; $H_m = 1$ gauss; $\omega_m = 2\pi \times 1000$ rad/second; $T = 1.2°K$; $dH_0/dt \approx 1/10$ gauss/second (A), 1/5 gauss/second (B), 2/5 gauss/second (C), 1 gauss/second (D) and 2 gauss/second (E).

ened lines can be described by the Bloch equations only under very specialized conditions.

Concerning ii, it appears that conditions can be found under which an undistorted line can be observed, with a good signal-to-noise ratio, for *any* T_1. For short T_1's, nonsaturated conditions can be maintained. For very long values of T_1, slightly nonadiabatic conditions should be maintained, and a linear field sweep that is fast enough to avoid any traces of line splitting (see Fig. 45). This set of conditions is discussed in Section 6.12. The amplitude of the signal under these conditions depends upon the external variables H_m, ω_m, dH_0/dt and H_1, but very little upon T_1 and T_2, if they are sufficiently long.

Experiments to check this were carried out, and will be reported elsewhere. It is concluded that the "moderate nonadiabatic" case 11 is quite reliable in determination of relative intensities of different lines and fine details of line shapes.

IX. ACKNOWLEDGMENTS

This work was carried out under the supervision and with the constant help of G. Feher. In many of the experiments the writer was assisted by E. Gere. The writer also profited from discussions with W. Blumberg, and suggestions made by P. W. Anderson.

X. NOTATION

$a \equiv$ Hyperfine interaction (except in Appendix C).
$b \equiv$ Dipolar interaction between an electron and a nucleus.
$H \equiv$ Magnetic field in the z-direction at any instant.
$H_0' \equiv$ Magnetic field at the center of an inhomogeneously broadened line.
$H_0 \equiv$ Magnetic field at the center of a spin packet, or at the center of the field modulation cycle.
$H' \equiv$ The width (on the H scale) of a nonsaturated spin packet distribution.
$H_1 \equiv$ Component of rotating magnetic field, perpendicular to the z-direction.
$H_m \equiv$ Peak modulating magnetic field (in the z-direction).
$\Delta H \equiv$ Width of inhomogeneously broadened line (defined as the second moment of the line).
$dH/dt \equiv$ Instantaneous value of time derivative of the magnetic field.
$dH_0/dt \equiv$ Magnetic field sweep (assumed constant).

$h(H - H_0')$ ≡ Shape function of inhomogeneously broadened line, normalized to unity: $\int_{-\infty}^{\infty} h(H - H_0')\, dH = 1$.

$\mathbf{i}, \mathbf{j}, \mathbf{k}$ ≡ Unit vectors along the coordinate axes.

J ≡ Exchange integral.

k ≡ $(1/H_1)(dH/dt)$, assumed constant or slowly varying.

k ≡ Boltzman's constant.

N ≡ Number of field modulation cycles.

s_e ≡ Electron spin.

s_n ≡ Nuclear spin.

$\mathbf{S}$ ≡ Magnetization vector.

S_x, S_y, S_z ≡ Components of magnetization vector.

$|S|$ ≡ Absolute value of magnetization vector.

S_0 ≡ Equilibrium value of magnetization vector.

S_N ≡ Absolute value of magnetization vector after N field modulation cycles.

t ≡ Time.

T_1 ≡ Longitudinal relaxation time.

T_2 ≡ Transverse relaxation time.

T ≡ Period of modulating field.

T ≡ Temperature.

u, w, p, q, x ≡ Dimensionless integration variables.

x, y, z ≡ Cartesian coordinates in space, or $z = x + iy$ coordinates in the complex plane.

α ≡ Fractional loss of magnetization during rapid passage (except in Appendix E).

γ ≡ Gyromagnetic ratio.

δ ≡ $H - H_0/H_1$.

ϵ ≡ $\mu_e H/kT$.

θ, φ ≡ Spherical coordinates.

μ_e ≡ Magnetic moment of electron.

μ_n ≡ Magnetic moment of nucleus.

χ' ≡ $S_x/2H_1$.

χ_f' ≡ Component of χ' at magnetic field modulation frequency.

χ'' ≡ $S_y/2H_1$.

χ_f'' ≡ Component of χ'' at magnetic field modulation frequency.

ω ≡ Microwave angular frequency.

ω_m ≡ Angular frequency of magnetic field modulation.

$\Delta\omega$ ≡ Difference in resonant angular frequencies of two electrons.

$\Delta\omega_{sys}$ ≡ Bandwidth of system.

APPENDIX A

χ' and χ'' for a Single Spin Packet, for Long Relaxation Times

A.1 *Approximations for Long Relaxation Times*

The Bloch equations are three equations in three variables, S_x, S_y, S_z and thus are difficult to solve. They can be solved relatively easily, if one variable can be eliminated. The "conventional" method is to eliminate S_z. For a sudden passage, we let S_z be a constant and for an adiabatic passage: $S_z = (\delta/\sqrt{1+\delta^2})S_0$, say. Once we know S_z, then S_x and S_y can be combined to one complex variable $S_x + iS_y$, and we have one linear first-order differential equation with one complex variable, which is integrable by quadratures.†

However, in a region intermediate between adiabatic and sudden passages, particularly for the transient solution, this method does not work.

If T_1, T_2 are long, the Bloch equations are in the rotating frame,

$$\frac{\partial S}{\partial t} = -\gamma H_1 \sqrt{1+\delta^2}\left[\frac{\mathbf{i}+\delta\mathbf{k}}{\sqrt{1+\delta^2}} \times \mathbf{S}\right];$$

i.e., $|S|$ is a constant and S moves on the surface of a sphere. Thus, S can be described by *two* variables — say, the spherical coordinates θ and φ, and the third variable is eliminated. But if we write the Bloch equations in terms of spherical coordinates, they are practically unmanageable.

However, a *portion* of the surface of a sphere can be covered quite accurately by a rectangular coordinate system. In the sudden case, S is limited to a region in the vicinity of the "north pole." In the adiabatic case, it is limited to a narrow strip from the north pole to the south pole along the $\varphi = 0°$ longitude. Even if $\gamma H_1^2 = dH/dt$, S will be in the "northern" hemisphere, which can reasonably well be covered by a cartesian net.

Thus, in any case the relevant portions of the sphere can be covered by a rectangular net. Using these coordinates, the Bloch equations are *two* linear first-order equations in two variables, which can be combined to *one* in one complex variable, which is soluble by quadratures. This solution, though only approximate for $\gamma H_1^2 \approx dH/dt$ does not break down for any ratio of these parameters. Of course, in the adiabatic and sudden cases it reduces to the well-known solutions.

In the intermediate case, S_x, S_y, S_z can be written down explicitly in terms of a relatively simple integral for which there seems to be no

† A more detailed discussion of this point is given by Salpeter.[4]

tabulated functions, but which may be evaluated numerically quite easily.

The Bloch equations are:

$$\frac{\partial S}{\partial t} = -\gamma H_1 \delta(\mathbf{k} \times \mathbf{S}) - \gamma H_1(\mathbf{i} \times \mathbf{S})$$

$$= -\gamma H_1 \sqrt{1+\delta^2}\left(\frac{\mathbf{i}+\delta\mathbf{k}}{\sqrt{1+\delta^2}} \times S\right)$$

(H_0 in z-axis direction, H_1 in x-axis direction). So, $|S| =$ constant $= S_0$ and $\mathbf{S}$ moves on the surface of a sphere (see Fig. 46); $\partial\mathbf{S}/\partial t$ is perpendicular to $\mathbf{r}$, where

$$\mathbf{r} = \frac{\mathbf{S}}{S_0} - \frac{\mathbf{i}+\delta\mathbf{k}}{\sqrt{1+\delta^2}},$$

and is of magnitude $\gamma H_1 \sqrt{1+\delta^2}\,|\mathbf{r}|$. Note that $\cot\theta = -\delta$ (θ = polar angle). Now, in case the adiabatic condition is satisfied, S will move along the $\varphi = 0$ longitude. While, if $\gamma H_1^2 \ll dH/dt$, ("sudden" case), S will remain close to the "north pole." In both cases, a *cartesian* coordinate system can cover those portions of the sphere in which S and the effective field move, and the metric will be closely Euclidean, since the

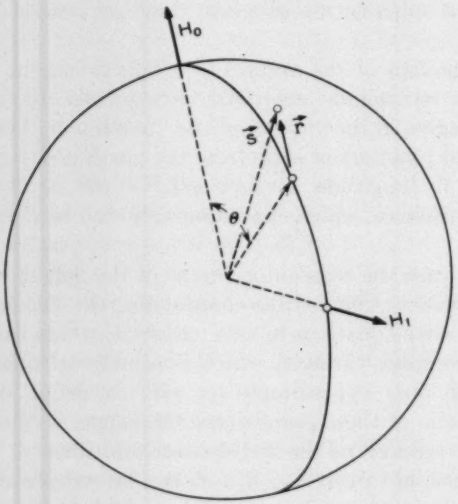


Fig. 46 — Position of magnetization vector in space.

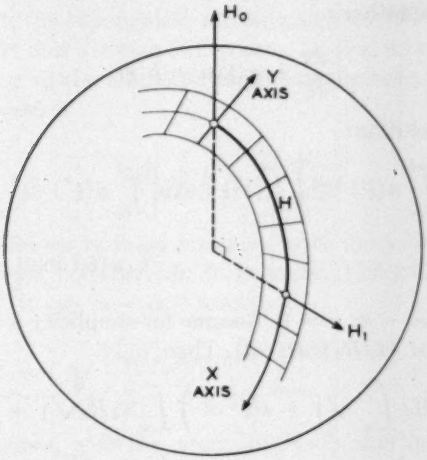


Fig. 47 — Almost-cartesian coordinate net covering a part of a sphere.

regions are "narrow" (see Fig. 47). In this coordinate system,

$$\frac{dx}{dt} = \gamma H_1 \sqrt{1 + \delta^2}\, y,$$

$$\frac{dy}{dt} = -\gamma H_1 \sqrt{1 + \delta^2}\,(x - x_0) \qquad (x_0 = \theta)$$

(see Fig. 48). Now the xy plane can be considered as the complex z plane, with $z = x + iy$. In it,

$$\frac{dz}{dt} = -i\gamma H_1 \sqrt{1 + \delta^2}\, z + i\gamma H_1 \sqrt{1 + \delta^2}\, x_0.$$

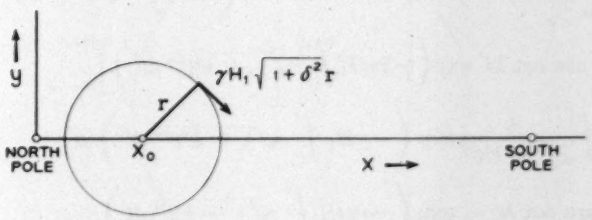


Fig. 48 — Diagram illustrating the differential equation satisfied (approximately) by the magnetization vector.

The differential equation

$$\frac{du}{d\xi} + a(\xi)u = b(\xi)$$

has the general solution:

$$u(\xi) = \exp\left[-\int_{\xi_0}^{\xi} a(\xi')\,d\xi'\right]\int_{\xi_0}^{\xi} b(\xi')\exp\left[\int_{\xi_0}^{\xi} a(\xi'')\,d\xi''\right]d\xi' + u(\xi_0)\exp\left[-\int_{\xi_0}^{\xi} a(\xi')\,d\xi'\right].$$

In our case, at $t = -\infty$, $z = 0$. Assume for simplicity a linear change in δ: $\delta = kt$, with $k = (1/H_1)(dH/dt)$. Then,

$$z = -\exp\left(-i\gamma H_1\int_{-\infty}^{t}\sqrt{1+k^2t'^2}\,dt'\right)\int_{-\infty}^{t} i\gamma H_1\sqrt{1+k^2t'^2}\,\text{arc cot } kt' \cdot\exp\left(i\gamma H_1\int_{-\infty}^{t'}\sqrt{1+k^2t''^2}\,dt''\right)dt'.$$

Since $x_0 = \theta = -arc\ \cot\ \delta$, where the branch of the *arc* cot function used throughout this calculation is the one for which *arc* cot $(-\infty) = 0$, then *arc* cot $(\infty) = -\pi$.

The "susceptibilities" are:

$$\chi' = \frac{1}{2}\frac{\omega\chi_0}{\gamma H_1}\sin(\text{Re } z), \qquad \chi'' = \frac{1}{2}\frac{\omega\chi_0}{\gamma H_1}\text{Im } z$$

in our approximation [(Im z) $\ll 1$, if the sphere is the unit sphere].

Integrating by parts,

$$\begin{aligned} z &= -\exp\left(-i\gamma H_1\int^{t}\sqrt{1+k^2t'^2}\,dt'\right) \\ &\quad\cdot\int_{-\infty}^{t} i\gamma H_1\sqrt{1+k^2t'^2}\,\text{arc cot } kt'\exp\left(i\gamma H_1\int^{t}\sqrt{1+k^2t''^2}\,dt''\right)dt' \\ &= -\left[\text{arc cot } kt'\exp\left(-i\gamma H_1\int_{t'}^{t}\sqrt{1+k^2t''^2}\,dt''\right)\right]_{t'=-\infty}^{t'=t} \\ &\quad -\int_{-\infty}^{t}\frac{k}{1+k^2t'^2}\exp\left(-\gamma H_1\int_{t'}^{t}\sqrt{1+k^2t''^2}\,dt''\right)dt' \\ &= -\text{arc cot } kt - \exp\left(-i\gamma H_1\int^{t}\sqrt{1+k^2t'^2}\,dt'\right) \\ &\quad\cdot\int_{-\infty}^{t}\frac{k}{1+k^2t'^2}\exp\left(i\gamma H_1\int^{t'}\sqrt{1+k^2t''^2}\,dt''\right)dt'. \end{aligned}$$

If $k/(\gamma H_1) \ll 1$, the exponential will be roughly $e^{i\gamma H_1 t'}$, and the integral will oscillate fast and average out to zero, leaving us with $z \approx -arc \cot kt$, corresponding to the adiabatic rapid passage case.

The sudden case is

$$\gamma H_1 \ll k \equiv \frac{1}{H_1}\frac{dH_0}{dt}.$$

If $\gamma H_1 \ll k$, kt must be large compared with unity before the exponential starts differing considerably from unity. Using the approximation $1 + k^2t^2 \approx k^2t^2$, it can be easily seen that

$$z(t) = i\sqrt{\frac{2\gamma H_1}{k}}\, e^{i\gamma H_1 k(t^2/2)} \int_{-\infty}^{\sqrt{(\gamma H_1 k)/2}\, t} e^{-iw^2}\, dw$$

This result agrees, with the results of Jacobsohn and Wangness[3] and Salpeter.[4]

Physically, **S** initially starts out from the north pole along the $\varphi = 0$ meridan, and eventually turns around in a circle of radius

$$\sqrt{\frac{\pi 2\gamma H_1}{k}} = \sqrt{\frac{2\pi\gamma H_1^2}{\frac{dH}{dt}}},$$

i.e., one that is smaller as the sudden approximation works better (see Figs. 10 and 49).

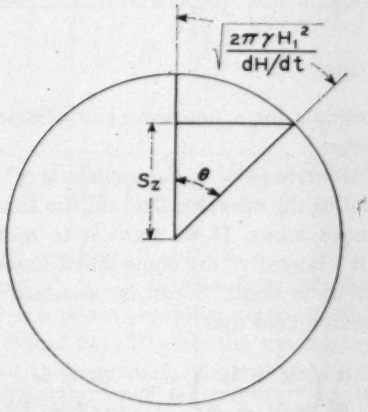


Fig. 49 — The loss of magnetization during an almost sudden passage.

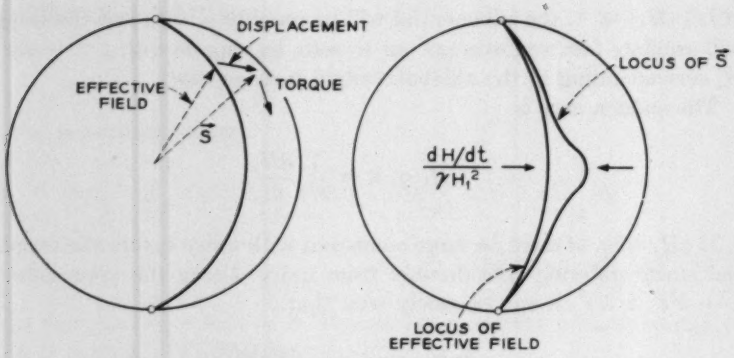


Fig. 50 — The locus of the magnetization vector for an almost adiabatic rapid passage.

If $\sqrt{(\pi 2\gamma H_1)/k} \ll 1$ for the movement along that circle,

$$S_z = S_0 \cos\theta \approx S_0\left(1 - \frac{\theta^2}{2}\right) = S_0\left(1 - \frac{\pi\gamma H_1}{k}\right);$$

i.e., the fractional loss of magnetization is $(\pi\gamma H_1^2)/(dH/dt)$, again a well-known result.

Note that the precession in the rotating frame taking place for large t is actually a rotation at a rate ω_0 in the fixed frame, i.e., the free induction tail. However, because the amplitude of our microwave signal at frequency ω is large compared with that of the free induction signal, one must consider the signal with respect to it, i.e., the beats of $\mathbf{S}$ with it.

A.2 *The Adiabatic Case*

The χ' and χ'' signals for a nonperfect adiabatic passage have been calculated by Salpeter.[4]

Physically, the occurrence of a nonvanishing χ'' can be seen as follows: If $\mathbf{S}$ is parallel to the effective field all the time, there is no torque on it. Thus, it cannot move. If we want it to move "longitudinally," we must displace it "laterally" by some small amount (see Fig. 50).

If the displacement is *small*, S can be assumed to move practically parallel to the effective field and

$$\frac{d\mathbf{S}}{dt} \approx \frac{d}{dt}\left(\frac{\mathbf{k} + \delta\mathbf{i}}{\sqrt{1+\delta^2}}\right) S = \frac{\mathbf{k} - \mathbf{i}\delta}{(\sqrt{1+\delta^2})}\,\dot{\delta}\,|\,S\,|.$$

Let $S = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. From the differential equation

$$\frac{d\mathbf{S}}{dt} = -\gamma H_1 \sqrt{1+\delta^2}\left(\frac{\mathbf{i}+\delta\mathbf{k}}{\sqrt{1+\delta^2}} \times \mathbf{S}\right)$$

and our approximation for $d\mathbf{S}/dt$, it follows that

$$\frac{\mathbf{k}-\mathbf{i}\delta}{(\sqrt{1+\delta^2})^3}\dot{\delta}\,|\,S\,| = -\gamma H_1(\delta \times \mathbf{j} - Sy\mathbf{i} + y\mathbf{k} - z\mathbf{j}),$$

$$y = -\frac{\dot{\delta}}{\delta H_1}\frac{1}{(1+\delta^2)^{\frac{1}{2}}}S,$$

$$\chi'' = \frac{\frac{dH}{dt}}{\gamma H_1^2}\frac{1}{\left[1+\left(\frac{H-H_0}{H_1}\right)^2\right]^{\frac{1}{2}}}\left(\tfrac{1}{2}\frac{\chi_0\omega}{\gamma H_1}\right).$$

This χ'' satisfies the conservation of energy theorem:

$$E = M_z H_0 = \frac{\chi_0\omega}{\gamma} H_0 \frac{\delta}{\sqrt{(1+\delta^2)^2}},$$

$$\frac{dE}{dt} = \chi_0 H_0^2 \frac{\dot{\delta}}{(1+\delta^2)^{\frac{1}{2}}} = 2H_1^2\omega\chi'',$$

$$\chi'' = \frac{\chi_0 H_0^2 \frac{1}{H_1}\frac{dH}{dt}}{\left[1+\left(\frac{H-H_0}{H_1}\right)^2\right]^{\frac{1}{2}}}\left(\frac{1}{2H_1^2\gamma H_0}\right)$$

$$= \frac{\frac{dH}{dt}}{\gamma H_1^2}\frac{1}{\left[1+\left(\frac{H-H_0}{H_1}\right)^2\right]^{\frac{1}{2}}}\left(\frac{1}{2}\frac{\chi_0\omega}{\gamma H_1}\right).$$

For χ' and χ'' signals, see Figs. 2 and 42.

Note that, although $\chi' > \chi''$, if phase-sensitive detection is used, χ' may average out by the phase sensitive detector while χ'' does not. It tends to produce a signal at 90° with the modulating field. However, any "burning out" of the magnetization will produce a χ' signal much stronger than χ'' (see Section 6.6).

The loss of S_z due to nonadiabaticity has been estimated, in the general case, by Zener[13] to be $e^{-2\pi\gamma'}$, where

$$\gamma' = \frac{{E_{12}}^2}{\hbar\left[\frac{d(E_1 - E_2)}{dt}\right]},$$

and E_1, E_2, E_{12} are defined by

$$\mathfrak{H}\Phi_1 = E_1\Phi_1 + E_{12}\Phi_2,$$

$$\mathfrak{H}\Phi_2 = E_{12}\Phi_1 + E_2\Phi_2,$$

where

Φ_1 = "spin up" wave function,

Φ_2 = "spin down" wave function, and

$\mathfrak{H}$ = the Hamiltonian of the system.

Here, $\mathfrak{H} = \hbar\gamma(H - H_0)S_z + \hbar\gamma H_1 S_x$ in the rotating frame. Thus, say, for $S = \frac{1}{2}$

$$E_1 = \hbar\gamma(H - H_0)S,$$

$$E_2 = \hbar\gamma(H_0 - H)S,$$

$$E_{12} = \hbar\gamma H_1 S.$$

Therefore,

$$\gamma' = \frac{\gamma {H_1}^2}{4\frac{dH}{dt}} \qquad (S = \tfrac{1}{2}),$$

and the fractional loss is

$$\exp\left(-\frac{\pi}{2}\frac{\gamma {H_1}^2}{\frac{dH}{dt}}\right).$$

A.3 *Conclusion*

Within the adiabatic region, the loss of magnetization due to nonadiabaticity is negligible.

APPENDIX B

χ' and χ'' for an Almost Sudden Passage of an Inhomogeneously Broadened Line, Employing a System with a Bandwidth Narrow Compared with the Line Width

The susceptibility of an individual spin packet for an almost sudden passage ($dH/dt \gg \gamma H_1^2$) is

$$\chi'(t') + i\chi''(t') = \chi_0 \frac{H_0}{2H_1} i \sqrt{\frac{2\gamma H_1}{k}} e^{i\gamma H_1 k(t^2/2)} \int_{-\infty}^{\sqrt{(\gamma H_1 k)/2}\, t'} e^{-iw^2}\, dw$$

as derived by Salpeter[4] (or see Appendix A). Here $k = (1/H_1)(dH/dt)$ is assumed to be constant, and it is assumed that the magnetic field sweeps through resonance at $t' = 0$.

For spin packets swept through resonance at a time t', the susceptibility at time t is

$$\chi'(t,t') + i\chi''(t,t') = \chi_0 \frac{H_0}{2H_1} i \sqrt{\frac{2\gamma H_1}{k}} \exp\left[i\gamma H_1 b \frac{(t - t'^2)}{2}\right] \int_{-\infty}^{\sqrt{(\gamma H_1 k)/2}(-t')} e^{-iw^2}\, dw.$$

At a time t' the magnetic field is:

$$H(t') = H(t) - \frac{dH}{dt}(t - t') = H(t) - H_1 k(t - t').$$

The susceptibility due to spin packets resonating at fields between $H - H_1 k(t - t')$ and $H - H_1 k(t - t' + dt')$ is

$$h[H - H_0' - H_1 k(t - t')] H_1 k\, dt' [\chi'(t,t') + i\chi''(t,t')].$$

Superimposing the signals due to the various packets,

$$\chi'(t) + i\chi''(t) = \int_{-\infty}^{\infty} dt'\, h[H - H_0' - H_1 k(t - t')] H_1 k \chi_0 \frac{H}{2H_1} \cdot i \sqrt{\frac{2\gamma H_1}{k}} \exp\left[i\gamma H_1 k \frac{(t - t')^2}{2}\right] \int_{-\infty}^{\sqrt{(\gamma H_1 k)/2}(t - t')} e^{-iw^2}\, dw.$$

Let $t'' = t - t'$ and $H = H_0' + H_1 kt$. Then,

$$\chi'(t) + i\chi''(t) = \int_{-\infty}^{\infty} dt''\, h[H_1 k(t - t'')] \frac{\chi_0 H_0'}{2} i\sqrt{2\gamma H_1 k} \cdot \exp\left(i\gamma H_1 k \frac{t''^2}{2}\right) \int_{-\infty}^{\sqrt{(\gamma H_1 k)/2}\, t''} e^{-iw^2}\, dw.$$

For a gaussian line,

$$h[H_1k(t-t'')] = \frac{1}{\sqrt{2\pi}\Delta H}\exp\left\{-\frac{[H_1k(t-t'')]^2}{2\Delta H^2}\right\}.$$

So that

$$\chi'(t)+i\chi''(t) = \frac{\chi_0 H_0'}{2\Delta H}\, i\sqrt{\frac{\gamma H_1 k}{\pi}}\int_{-\infty}^{\infty} dt''\exp\left\{-\frac{[H_1k(t-t'')]^2}{2\Delta H^2}\right\}$$
$$\cdot\exp\left(i\gamma H_1 k\,\frac{t''^2}{2}\right)\int_{-\infty}^{\sqrt{(\gamma H_1 k/2)}\,t''} e^{-iw^2}\,dw.$$

Let

$$\sqrt{\frac{\gamma H_1 k}{2}}\,t'' = u, \qquad \sqrt{\frac{\gamma H_1 k}{2}}\,t = u_0.$$

Then

$$\chi'(t)+i\chi''(t) = \frac{\chi_0 H_0'}{2\Delta H}\, i\sqrt{\frac{\gamma H k_1}{\pi}}\sqrt{\frac{2}{\gamma H_1 k}}$$
$$\cdot\int_{-\infty}^{\infty} du\exp\left[-\frac{H_1 k}{\gamma\Delta H^2}(u-u_0)^2\right]e^{iu^2}\int_{-\infty}^{u} dw\, e^{-iw^2}.$$

If $H_1k/\gamma\Delta H^2 \ll 1$, the first term in the integrand may be assumed to be unity, and the integral is

$$\int_{-\infty}^{\infty} du\, e^{iu^2}\int_{-\infty}^{u} dw\, e^{-iw^2}.$$

This integral is the integral of $e^{i(u^2-w^2)}$ over the half of the u–w plane for which $u > w$ (see Fig. 51).

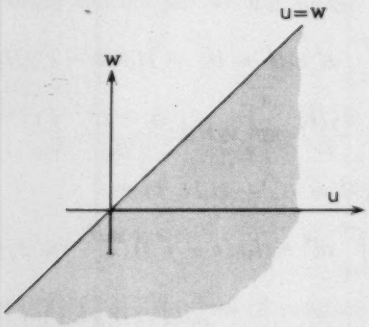


Fig. 51 — The half-plane over which integration of Appendix B is carried out.

This integral equals

$$\int_{-\infty}^{\infty} dw \int_{w}^{\infty} du\, e^{i(u^2-w^2)}.$$

In this integral, u and w may be renamed so that

$$\int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dw\, e^{i(u^2-w^2)} = \int_{-\infty}^{\infty} du \int_{u}^{\infty} dw\, e^{i(w^2-u^2)}.$$

Therefore

$$(\chi' + i\chi'') - (\chi' + i\chi'')^* = 2i\chi'' = \frac{\chi_0 H_0'}{\sqrt{2\pi}\Delta H}\, i \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dw\, e^{i(u^2-w^2)},$$

$$\int_{-\infty}^{\infty} e^{iu^2}\, du = (1+i)\sqrt{\frac{\pi}{2}},$$

$$\int e^{-iw^2}\, dw = (1-i)\sqrt{\frac{\pi}{2}}.$$

Thus,

$$\chi'' = \frac{\pi}{2}\frac{\chi_0 H_0'}{\sqrt{2\pi}\Delta H},$$

for an "infinitely" wide line. For a finite line, we evidently have

$$\chi'' = \frac{\pi}{2}\chi_0 H h(H - H_0')$$

approximately, where $h(H - H_0')$ is the line shape, normalized to unity, and the approximation is certainly good if $dH/dt \ll \gamma\Delta H^2$.

When the direction of sweep is reversed, χ'' does not change sign, and it is independent of dH/dt and H_1. [However, the area, or integrated signal, is proportional to $1/(dH/dt)$.]

For an "infinitely" wide line, χ' is zero, since the equation

$$\int du \int dw\, e^{i(u^2-w^2)} = \int du \int dw\, e^{i(w^2-u^2)}$$

indicates that the imaginary component of this integral vanishes, and χ' is proportional to this component. Thus, the χ' "wiggles" tend to *cancel* each other. A nonvanishing χ' is due to the fact that the line is not infinitely wide.

As derived before, for a gaussian line

$$\chi'(t) + i\chi''(t) = \frac{\chi_0 H}{\sqrt{2\pi}\Delta H}\, i \int_{-\infty}^{\infty} du \exp\left[-\frac{H_1 k}{\gamma \Delta H^2}(u - u_0)^2\right] \cdot e^{iu^2} \int_{-\infty}^{u} e^{-iw^2}\, dw.$$

Differentiating with respect to t, at $t = 0$

$$\frac{d}{dt}[\chi'(0) + i\chi''(0)] = \frac{\chi_0 H}{\sqrt{2\pi}\Delta H}\, i \frac{2H_1 k}{\gamma \Delta H^2} \sqrt{\frac{\gamma H_1 k}{2}} \cdot \int_{-\infty}^{\infty} du\, u \exp\left(-\frac{H_1 k}{\gamma \Delta H^2} u^2\right) e^{iu^2} \int_{-\infty}^{u} e^{-iw^2}\, dw.$$

Now,

$$\begin{aligned}
&\int_{-\infty}^{\infty} du \int_{-\infty}^{u} dw\, u \exp\left(-\frac{H_1 k}{\gamma \Delta H^2} u^2 + iu^2 - iw^2\right) \\
&\quad = \int_{-\infty}^{\infty} dw \int_{w}^{\infty} du\, u \exp\left(-\frac{H_1 k}{\gamma \Delta H^2} u^2 + iu^2 - iw^2\right) \\
&\quad = \frac{-1}{2\left(i - \frac{H_1 k}{\gamma \Delta H^2}\right)} \int_{-\infty}^{\infty} dw \exp\left(-\frac{H_1 k}{\gamma \Delta H^2} w^2 + iw^2 - iw^2\right) \\
&\quad = \frac{-1}{2\left(i - \frac{H_1 k}{\gamma \Delta H^2}\right)} \sqrt{\frac{\pi \gamma \Delta H^2}{H_1 k}}.
\end{aligned}$$

If $dH/dt \ll \gamma \Delta H^2$, $H_1 k/\gamma \Delta H^2$ can be neglected when compared with unity, and we have

$$\left(\frac{d\chi'}{dt}\right)_{H=H_0'} = -\frac{\chi_0 H}{\sqrt{2\pi}\Delta H} \frac{2dH/dt}{\gamma \Delta H^2} \sqrt{\frac{\gamma dH/dt}{2}} \left(\frac{1}{2}\sqrt{\frac{\pi \gamma \Delta H^2}{dH/dt}}\right)$$

or, since $H = H_0 + (dH/dt)t$,

$$\left(\frac{d\chi'}{dH}\right)_{H=H_0'} = -\frac{H_0 \chi_0}{2\Delta H} \frac{1}{\Delta H}.$$

At $H = H_0' + \Delta H$,

$$\chi' \approx \chi'(H_0') + \left(\frac{d\chi'}{dH}\right)_{H=H_0'} \Delta H = -\frac{\chi_0 H}{2\Delta H}$$

$$= -\frac{1}{\sqrt{2\pi}} \chi_0 H h(H_0' + \Delta H),$$

which is of the same order as χ''.

We have calculated χ'' only at the center (of a very wide) line, and, as to χ', only $d\chi'/dH$ has been calculated at the center of a gaussian line. However, for a microwave spectrometer the bandwidth of which is *small* compared with the line width, these results can be used to calculate χ' and χ'' for any part on any line, since the higher-order terms are due to spin packets off resonance, and these tend to produce high-frequency beats with the microwave signal. If the bandwidth of the system is small, these beats are not observed.

Let

$$\chi'(H) = ah(H - H_0') + b\frac{dh(H - H_0')}{dH} + c\frac{d^2h(H - H_0')}{dH^2} + \cdots,$$

and introduce a similar expression for χ''.

The first term in the above expression is the susceptibility of an infinitely wide line. This vanishes for χ', and has been calculated for χ''. Thus, the second term must be considered in the expression for χ', but not in that for χ''. So, let

$$\chi' = b\left[\frac{dh(H - H_0')}{dH}\right],$$

Then

$$\left(\frac{d\chi'}{dH}\right)_{H=H_0'} = b\left(\frac{d^2h(H - H_0')}{dH^2}\right)_{H_0'} = b\frac{1}{\sqrt{2\pi}\Delta H}\frac{1}{\Delta H^2}$$

at the center of a gaussian line. But, in this case,

$$\frac{d\chi'}{dH} = -\frac{\chi_0 H}{2\Delta H}\frac{1}{\Delta H},$$

so

$$b = -\Delta H\chi_0 H\sqrt{\frac{\pi}{2}},$$

$$\chi' = -\sqrt{\frac{\pi}{2}}\chi_0 H\frac{dh(H - H_0')}{dH}\Delta H,$$

$$\chi'_{\max} = \tfrac{1}{2}\chi_0\frac{H}{\Delta H}\frac{1}{\sqrt{e}},$$

while

$$\chi''_{\max} = \tfrac{1}{2}\chi_0\frac{H}{\Delta H}\sqrt{\frac{\pi}{2}}.$$

The difference between electron spin resonance and nuclear magnetic resonance may be noted. (The NMR case has been treated by Gabillard.[14])

With common NMR systems, the bandwidth may be large compared with the line widths, and the "beats" far off the line can be observed — and they yield the Fourier transform of the line. {Since the superposition $\int h(H - H_0') \exp [i\gamma(H - H_0')t]d(\gamma H)$ is essentially a Fourier transform of $h(H - H_0')$}. This is the signal observed in region III (see Section IV). The signal of region I is at the microwave frequency, but weak.

The agreement between the above calculations and experiment is excellent.

Fig. 12 confirms the following features:

i. *Shapes* of the signals are "dispersion" for χ', "absorption" for χ''.
ii. χ' and χ'' are of about the same strength.
iii. The signals are independent of dH_0/dt.

APPENDIX C

χ' and χ'' for a Nonadiabatic Passage, Nonuniform Rate of Sweep

C.1 χ' and χ'' for a Triangular Wave and Almost Sudden Transition of an Inhomogeneously Broadened Line

The general solution of the first differential equation of Appendix A is

$$z(t'') = \exp\left(i\frac{\gamma H_1 k}{2}t''^2\right)\int_{t_0}^{t''} i\gamma H_1 S_z \exp\left(-i\frac{\gamma H_1 k}{2}t'^2\right)dt' + \exp\left[i\frac{\gamma H_1 k}{2}(t''^2 - t_0^2)\right]z(t_0).$$

This solution corresponds to the signal from a spin packet swept through resonance at $t'' = 0$. t_0 may be any arbitrary time. Let

$$H = \begin{cases} H_0 + H_1 kt & t < 0, \\ H_0 - H_1 kt & t > 0. \end{cases}$$

At $t = 0$, the signal due to a packet swept through at $t = -t_0$ is

$$z(0, -t_0) = \exp\left(i\frac{\gamma H_1 k}{2}t_0^2\right)\int_{-\infty}^{t_0} i\gamma H_1 S_z \exp\left(-i\frac{\gamma H_1 k}{2}t''^2\right)dt''.$$

This packet will again resonate at $t = t_0$. For $t > 0$,

$$z(t) = \exp\left[-i\frac{\gamma H_1 k}{2}(t-t_0)^2\right] i\gamma H_1 \int_0^t S_z \exp\left[i\frac{\gamma H_1 k}{2}(t'-t_0)^2\right] dt'$$
$$+ \exp\left[-i\frac{\gamma H_1 k}{2} t(t-2t_0)\right] \exp\left(i\frac{\gamma H k}{2} t_0^2\right)$$
$$\cdot \int_{-\infty}^{t_0} i\gamma H_1 S_z \exp\left(-i\frac{\gamma H_1 k}{2} t'^2\right) dt'.$$

Superimposing the signals due to the various spin packets, as in Appendix B, we have, assuming $S_z = S_0$,

$$\chi'(t) + i\chi''(t) = \int_{-\infty}^{\infty} dt_0 h(H(t) - H_0' - H_1 k(t-t_0)) H_1 k \chi_0 \frac{H}{2H_1}$$
$$\cdot i\gamma H_1 \left(\exp\left[-i\frac{\gamma H_1 k}{2}(t-t_0)^2\right] \int_0^t \exp\left[i\frac{\gamma H_1 k}{2}(t'-t_0)^2\right] dt'\right.$$
$$\left. + \exp\left\{-i\frac{\gamma H_1 k}{2}[t(t-2t_0) - t_0^2]\right\} \int_{-\infty}^{t_0} \exp\left(-i\frac{\gamma H_1 k}{2} t'^2\right) dt'\right).$$

For an "infinitely" wide line, let $h(H - H_0') = 1/(\sqrt{2\pi}\Delta H)$:

$$\chi'(t) + i\chi''(t) = \frac{i}{\sqrt{2\pi}} \frac{H}{\Delta H} \chi_0 \left[\int_{-\infty}^{\infty} du\, e^{-iu^2} \int_{(u-\sqrt{(\gamma H_1 k)/2})t}^{u} e^{iw^2} dw\right.$$
$$\left. + e^{-\gamma H_1 k t^2} \int_{-\infty}^{\infty} du\, e^{+iu^2} \int_{-\infty}^{(u-\sqrt{(\gamma H_1 k)/2})t} e^{-iw^2} dw\right].$$

To evaluate the above integrals, we may employ the transformation

$$p = \frac{u+w}{\sqrt{2}}, \qquad q = \frac{u-w}{\sqrt{2}}.$$

Then

$$\frac{\partial(q,p)}{\partial(u,w)} = 1,$$

$$u^2 - w^2 = 2pq$$

and

$$\int_{-\infty}^{\infty} du \int_{(u-\sqrt{(\gamma H_1 k)/2})t}^{u} dw\, e^{-i(u^2-w^2)}$$

$$= \int_{-\infty}^{\infty} dp \int_{0}^{(\sqrt{(\gamma H_1 k)/2})t} dq\, e^{-2ipq}$$

$$= \int_{-\infty}^{\infty} dp \left(-\frac{1}{2ip}\right) (\cos p\sqrt{\gamma H_1 k}\, t - i \sin p\sqrt{\gamma H_1 k}\, t - 1)$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} dp\, \frac{\sin \sqrt{\gamma H_1 k}\, pt}{p} = \begin{cases} \frac{\pi}{2} & \text{for} \quad t = 0 \\ 0 & \text{for} \quad t = 0, \dagger \end{cases}$$

$$\int_{-\infty}^{\infty} du \int_{-\infty}^{u-\sqrt{(\gamma H_1 k)/2}t} dw\, e^{i(u^2-w^2)}\, dw$$

$$= \int_{-\infty}^{\infty} dp \int_{\sqrt{(\gamma H_1 k)/2}t}^{\infty} dq\, e^{2ipq}$$

$$= \int_{-\infty}^{\infty} dp\, \frac{1}{2ip}\, (e^{i\infty} - e^{i\sqrt{\gamma H_1 k}\, pt})$$

$$= -\frac{1}{2}\int_{-\infty}^{\infty} dp\, \frac{\sin \sqrt{\gamma H_1 k}\, pt}{p} = -\frac{\pi}{2}, \ddagger$$

$$\chi'(t) + i\chi''(t) = \frac{i}{\sqrt{2\pi}}\frac{H}{\Delta H}\chi_0 \frac{\pi}{2}(1 - e^{i\gamma H_1 k t^2}),$$

$$\chi''(H) = \frac{\pi}{2}\chi_0 H h(H - H_0')(1 - \cos \gamma H_1 k t^2),$$

$$\chi'(H) = \frac{\pi}{2}\chi_0 H h(H - H_0') \sin \gamma H_1 k t^2.$$

The assumptions are the same as those discussed in Appendix B.

† The integral

$$\int_{-\infty}^{\infty} du\, e^{-iu^2} \int_{u-\sqrt{(\gamma H_1 k)/2i}}^{u} dw\, e^{iw^2}$$

seems to be a discontinuous function of t. It obviously vanishes for $t = 0$, and equals $\pi/2$ for $t < 0$. This may be due to the nonphysical assumption of an infinite linewidth. However, practically, dH/dt cannot change abruptly. For a sinusoidal sweep, with its smooth change of dH/dt, no discontinuity is expected even for an infinitely wide line.

‡ The term $e^{i\infty}$ presents no difficulties. Mathematically, it can be removed by a slight generalization of the concept of convergence. Physically it can be removed by considering any amount of relaxation.

We have a steady χ'' equal to that due to a sweep at a constant rate, and additional oscillations at a frequency $\omega' = 2\gamma H_1 kt = 2\gamma(dH/dt)t$. (At $t = 0$, H is at the peak of the triangular waveform.) The χ' and χ'' to be expected are shown in Fig. 52.

The experimental trace on Fig. 53(a) resembles the expected trace. However, the magnetization under the experimental conditions was not uniform, so a very close agreement between the preceding theory and this experiment cannot be expected.

c.2 *Some Additional Considerations Concerning Case 6.11*

Consider the extreme nonadiabatic case ($\gamma H_1^2 \lll \omega_m H_m$). Let the width of the "unburnt" region of spins, [within the range $H_0 + (dH_0/dt)\ t \pm H_m$, see Fig. 38] be H'. Then, we may consider the following cases:

i. $H' < H_1$; i.e., all spin packets that are more than H_1 away from the extremum of the modulation cycles are saturated and "destroyed."

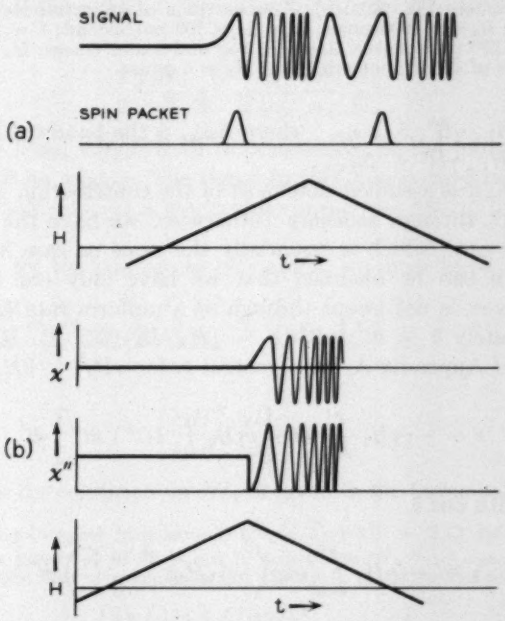


Fig. 52 — Illustration of the signal resulting from a triangular field modulation (almost sudden passage): (a) individual spin packet; (b) inhomogeneously broadened line.

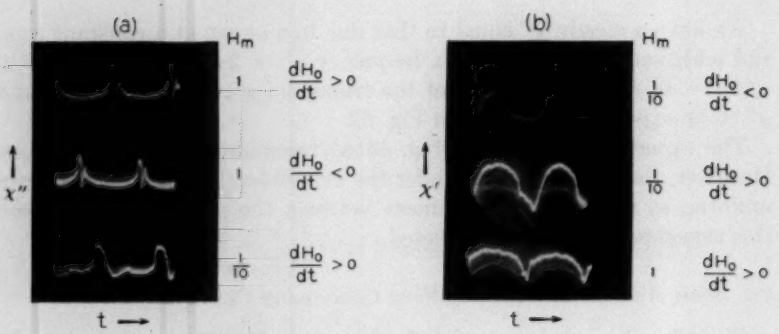


Fig. 53 — (a) χ'' scope trace of an extremely nonadiabatic passage. Data: $H_1 \approx 1/1000$ gauss; $\omega_m = 2\pi \times 100$ rad/second; $T = 1.2°K$; concentration, 6.5×10^{16} phosphorus atoms/cc. Top and center traces, $H_m = 1$ gauss, opposite senses of dH_0/dt; bottom trace, $H_m = 1/10$ gauss. The "wiggles" are probably damped at a rate determined by the T_2 of the sample, which is roughly of order 10^{-4} second (see Gordon and Bowers[15]); note, however, that the average of the 'wiggles' does not vanish, but seems to be proportional to $|S_x|$ (this is in agreement with the calculations of Appendix C); note also the sudden drop of χ'' when $dH/dt = 0$, predicted there (see Fig. 52). (b) χ' scope trace of an extremely nonadiabatic passage. Data: $H_1 = 1/1000$ gauss; $\omega_m = 2\pi \times 100$ rad/second; $T = 1.2°K$; concentration, 6.5×10^{16} phosphorus atom/cc. Top and center traces: $H_m = 1/10$ gauss, opposite senses of dH/dt; bottom trace: $H_m = 1$ gauss.

ii. $H' \gg H_1$, $\gamma H' > \Delta\omega_{sys}$, where $\Delta\omega_{sys}$ is the bandwidth of system.

iii. $H' \gg H_1$, $\gamma H' < \Delta\omega_{sys}$.

In all cases, it is assumed that most of the contribution is due to spin packets swept through *suddenly* (otherwise, we have the "moderate" nonadiabatic case, which is essentially the same as case 8).

In case i it can be assumed that we have only one spin packet, which, however, is not swept through at a uniform rate ($\delta = kt$), but at approximately $\delta = at^2$, with $a = (H_m/H_1)(\omega_m^2/2)$. Repeating the calculation of Appendix A for a general δ, for $\gamma H_1^2 \ll dH/dt$, we have

$$z = -i\gamma H_1 \int_{-\infty}^{t} \exp\left[i\gamma H_1 \int_{t'}^{t} \delta(t'')\, dt'' \right] dt'$$

as before. With our δ,

$$z \approx -i\gamma H_1 \int_{-\infty}^{t} \exp\left(-i\gamma H_1 a\, \frac{t^3 - t'^3}{3} \right) dt' .$$

Let

$$w \equiv \frac{a\gamma H_1}{3}\, t'^3 .$$

Then

$$z(t) = -i\sqrt[3]{\frac{(3\gamma H_1)^2}{a}}\exp\left(-i\frac{a\gamma H_1}{3}t^3\right)\int_{-\infty}^{\sqrt[3]{(a\gamma H_1)/3}}\frac{e^{iw}}{w^{\frac{1}{3}}}\,dw$$

(the integral can, if desired, be expressed in terms of the incomplete Γ function of a complex argument). At $t = 0$,

$$z(t) = -i\sqrt[3]{\frac{(\gamma H_1)^2}{9\dfrac{H_m}{H_1}\dfrac{\omega_m^2}{2}}}\int_{-\infty}^{0}\frac{e^{iw}}{w^{\frac{1}{3}}}\,dw,$$

and the amount burnt per cycle is

$$2\left(\frac{2}{9}\frac{\gamma H_1^2}{H_m\omega_m}\frac{\gamma H_1}{\omega_m}\right)^{\frac{1}{3}}\left(\int_0^\infty\frac{\cos w}{w^{\frac{1}{3}}}\,dw\right)^2.\ \dagger$$

Due to the slow rate of change of H near the extrema of the modulating cycle, there is a strong burnout, although $(\gamma H_1^2)/(H_m\omega_m)$ is small.‡

The amplitude of the signal should be of the order of $z(0)$ multiplied by the width of the region — which is of the order H_1 ; that is, the signal is roughly

$$\frac{1}{2}\frac{\chi_0 H_0}{\Delta H}\sqrt[3]{\frac{\gamma H_1^2}{H_m\omega_m}\frac{\gamma H_1}{\omega_m}\frac{H_1}{H_m}}.$$

Actually, for lower dH/dt than that indicated in the footnote ‡ below, the signal will be weaker. The signal in this case is very small, and this case is very inefficient. It is worthwhile to *increase* dH_0/dt to get a bigger signal.

Note that, for the passage to be sudden, we must have

$$\frac{dH}{dt} \approx H_m\omega_m\sqrt{2\frac{H'}{H_m}} \approx H_m\omega_m\sqrt{2\frac{H'}{H_m}} \gg \gamma H_1^2$$

that is,

$$\frac{\gamma H_1^2}{\omega_m H_m}\frac{\gamma H_1}{\omega_m} \ll 1$$

in addition to the condition on dH_0/dt given in the footnote‡.

† The definite integral here equals $(3\sqrt{3/2})(1/3!) = 2.32$ (see Jahnke and Emde[17]). For a power of 40 db below 10 milliwatts ($H_1 \approx \frac{1}{10000}$ gauss), 1000 cps, $H_m = 1$ gauss the loss of magnetization per cycle is approximately

$$2\left(\frac{2}{9}\right)\left(\frac{17.6}{6280}\right)\left(\frac{17600}{6280}\right)^{\frac{1}{3}}(2.32)^2 = \frac{1}{6}.$$

‡ For this case to apply under the above experimental conditions, dH_0/dt should not sweep more than H_1 in about 6 cycles; i.e., $dH_0/dt < (\omega_m/2\pi)(H_1)(\frac{1}{6})$ or $dH_0/dt < \frac{1}{6}$ gauss second.

In case ii, we may use the "superposition of wiggles" calculated in Appendix B, and use the estimate of Section IV for the burnout rate:

$$\alpha = \frac{\pi\gamma H_1^2}{\dfrac{dH}{dt}}, \qquad \frac{dH}{dt} \approx \frac{\pi}{2}\frac{H'}{H_m}\omega_m H_m$$

near the extrema of the modulation cycle. The distance that H progresses in $N/2$ modulation cycles is

$$H' = \frac{dH_0}{dt}\frac{2\pi}{\omega_m}\frac{N}{2}.$$

Thus,

$$\frac{dH}{dt} \approx \frac{\pi}{2}\frac{H'}{H_m}\omega_m H_m \approx \frac{\pi}{2}\frac{dH_0}{dt}\frac{2\pi}{\omega_m}\frac{N}{2}\omega_m = \frac{\pi^2}{2}N\frac{dH_0}{dt},$$

$$\alpha(N) = \frac{2\gamma H_1^2}{\pi N\dfrac{dH_0}{dt}}.$$

This is the loss of a spin packet during the Nth sweep through it. This formula holds for the case when dH_0/dt is not so small that, for $N = 1$, $H' < H_1$; i.e., this only holds for: $dH_0/dt > (\omega_m H_1)/\pi$ or, say, $dH_0/dt >$ 2 gauss/second at 40 db and 1000 cps. Otherwise, the loss during those cycles for which $H' < H_1$ should better be approximated as in case i.

The total loss is estimated as follows:

$$\ln \prod_N [1 - \alpha(N)] = \sum_N \ln [1 - \alpha(N)] \approx -\sum \alpha(N)$$

$$\approx -\frac{2\gamma H_1^2}{\pi dH_0/dt}\ln N$$

for small $\alpha(N)$. Thus

$$S_N \approx S_0 \exp\left(-\ln N\frac{2\gamma H_1^2}{\pi dH_0/dt}\right)$$

and

$$\frac{S_N}{S_0} = \frac{1}{e} \qquad \text{for} \quad N \approx \exp\left(\frac{\pi dH_0/dt}{2\gamma H_1^2}\right).$$

This width will be appreciable if $dH_0/dt > \gamma H_1^2$. (This condition is

usually more exacting than $dH_0/dt > (\omega_m/\pi)H_1$, since $\gamma H_1 > \omega_m$ except at very low powers.) For this N,

$$H' \approx \frac{dH_0}{dt}\left(\frac{2\pi}{\omega_m}\right)^{\frac{1}{2}} \exp\left(\frac{\pi dH_0/dt}{2\gamma H_1^2}\right),$$

and this gives the width of the nonburnt region.

Note that, if $dH_0/dt > \omega_m H_m$, the direction of sweep of field will not reverse itself at all. Then we shall have (if $dH_0/dt - \omega_m H_m \gg \gamma H_1^2$) very little burnout, and case 9 applies.

Here, T_2 has been assumed to be small compared with $2\pi/\omega_m$. At these low powers, $H_1 < H_{1oc}$, and the "zero field" value of T_2 can be employed.† For different T_2, see the discussion of Section 6.9.

This calculation gives $|S|$ for both cases ii and iii, essentially. The χ' and χ'' signals are calculated as in Section 6.8. (In case iii there will be some additional "wiggles".) Note that the expressions for χ' and χ'' apply not only for the natural line shape $h(H - H_0')$, but for any shape — such as the one we have here, which is determined by burning. This also implies that χ' and χ'' are of the same order of magnitude. However, the components at the field modulation frequency need not be of the same order of magnitude.

Note that when $dH/dt > 0$, and $dh(H - H_0)/dH > 0$, on the one hand, and $dH/dt < 0$, and $dh(H - H_0)/dH < 0$, on the other hand, the sign of χ' is the *same*. Thus, the scope traces will be as shown in Fig. 54 and the recorder traces as in Fig. 55. Experimental traces are shown in Figs. 42 and 53.

APPENDIX D

Superposition of Adiabatic Fast Passage Lines

For adiabatic fast passage,

$$\chi' \propto \frac{1}{\sqrt{1+\delta^2}} \qquad \delta = \frac{H - H_0}{H_1}.$$

Thus, for large δ, χ' falls down like $1/\delta$, and the superposition of signals due to packets of an infinitely broad line will yield an infinite result. Under actual physical conditions, there will be always something to limit χ': a finitely wide line, relaxation effects, etc. However, the calcu-

† This value has been measured by Gordon and Bower[15] and is approximately 10^{-4} second for the sample employed here.

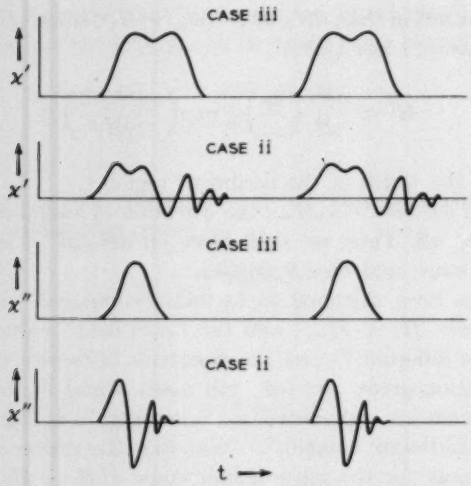


Fig. 54 — Expected χ' and χ'' scope traces of case 11; wiggles are observed when $\gamma H_1 \langle\langle\langle \omega_m H_m$, and $\omega_m T_2$ is not very small compared with unity.

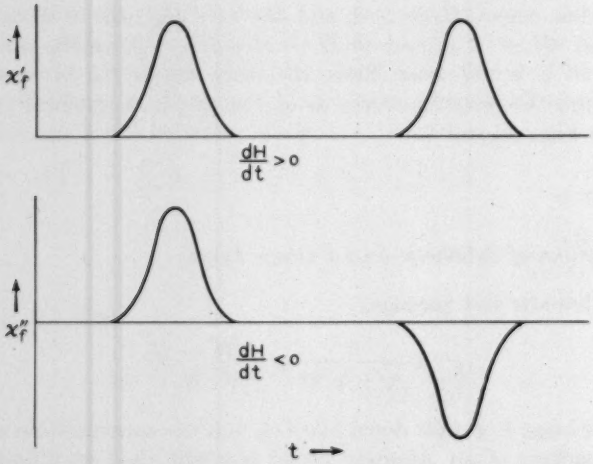


Fig. 55 — Expected recorder traces of case 11 (system relaxes between sweeps); the true line shape may be observed, sometimes even if H_m is *not* small compared with ΔH (the width of the nonburnt region must then be small). Compare with Figs. 42(a) and (b).

lation of the χ' signal involves integrals of the form

$$\int \frac{f(x)}{\sqrt{1+x^2}},$$

which are difficult to compute.

The case in which $f(x)$ is gaussian (i.e., a gaussian line, with relaxation neglected), has been treated by Feher[12] and Portis.[5] For it,

$$\int_{-\infty}^{\infty} \frac{h(H'-H_0)}{\sqrt{1+\delta^2}} \frac{dH'}{H_1} = \frac{1}{2\sqrt{2\pi}} \exp\left[-\left(\frac{H_1}{2\Delta H}\right)^2\right] K_0\left[\left(\frac{H_1}{2\Delta H}\right)^2\right]$$

for

$$h(H'-H_0) = \frac{1}{\sqrt{2\pi}\Delta H} \exp\left[-\frac{(H'-H_0)^2}{2\Delta H^2}\right],$$

and, for $H_1 \ll \Delta H$, this integral is, $(1/\sqrt{2\pi})$ log $(2\Delta H/H_1)$. This result will be of use for "plain" fast passage over a whole inhomogeneously broadened line, if

$$\frac{dH/dt}{\Delta H} \gg \frac{1}{\sqrt{T_1 T_2}}.$$

If field modulation is employed, this result does no longer apply (except in the "nonburnt" case 5). Practically, the width of the distribution of spins will be determined either by relaxation, or by the modulating field, H_m (or both).

The first case is difficult to treat. Portis[5] assumes an exponential distribution of packets,

$$\exp\left(-\left|\frac{H-H_0}{H_m\omega_m T_1}\right|\right),$$

where H is the magnetic field, H_0 is the resonant field of a packet and H_m is the peak modulating field.

This expression assumes that packets not yet swept over, are in a partly "relaxed" state given by this exponential; i.e., there is $t \rightarrow -t$ symmetry. As a matter of fact, the integral

$$\int \frac{e^{-|\alpha x|}}{\sqrt{1+x^2}}\, dx$$

is difficult to evaluate, so this approximation is not very helpful, practically.

The simplest approximation is, probably, to replace the actual $|S|$ distribution by a triangular one (see Fig. 56). In the case of an envelope

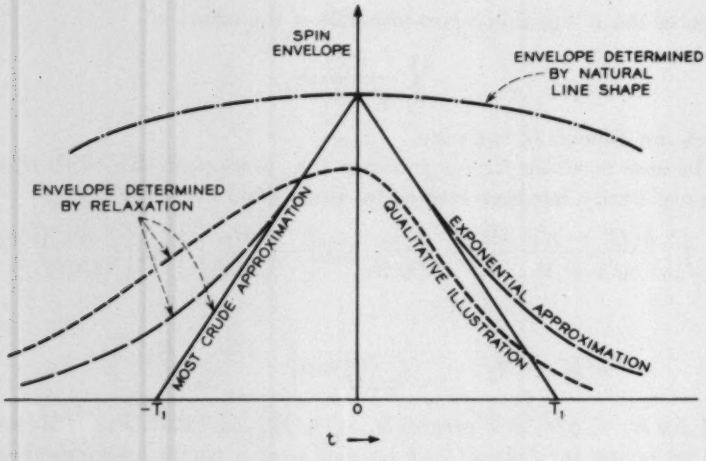


Fig. 56 — Envelope of the magnetization S_z, for an inhomogeneously broadened line under various experimental conditions.

determined by relaxation, it is reasonable to assume a width $t = T_1$, or

$$H' = \frac{dH}{dt} T_1 = \omega_m H_m T_1$$

in the center of the modulation cycle, for $T_1 = T_2$. (For $T_1 \neq T_2$, one might perhaps postulate a width $t = \sqrt{T_1 T_2}$.)

Thus,

$$\int_{-H_m\omega_m T_1}^{+H_m\omega_m T_1} \frac{1 - \left|\frac{H - H_0}{H_m\omega_m T_1}\right|}{\sqrt{1 + \left(\frac{H - H_0}{H_1}\right)^2}} \frac{d(H - H_0)}{H_1}$$

$$= 2 \log \frac{H_m\omega_m T_1 + \sqrt{H_1 + (H_m\omega_m T_1)^2}}{H_1}$$

$$- 2 \frac{1}{H_m\omega_m T_1} \left(\sqrt{H_1^2 + (H_m\omega_m T_1)^2} - H_1\right).$$

For adiabatic rapid passage, $H_1/(H_m\omega_m) \ll T_1$, so the integral is

$$\approx 2 \log \frac{2H_m\omega_m T_1}{H_1} - \left(\frac{H_1}{H_m\omega_m T_1}\right)^2 \approx 2 \ln \frac{2H_m\omega_m T_1}{H_1}.$$

Actually, since the integral $\int dx/\sqrt{1 + x^2}$ diverges only logarithmi-

cally, different shapes will change only the factor "2" in the logarithm, which has little effect on the amplitude and probably cannot be checked experimentally except with high-precision equipment.

If the width of the distribution is limited by H_m, it is reasonable to assume that the packets *outside* the interval $H_0 - H_m < H < H_0 + H_m$ do not yield any signal, particularly if $H_1 \ll H_m$.

APPENDIX E

Second-Order Effects Occurring When a Bridge Is Tuned to Observe χ' or χ'' Signals

When a "bridge" system is used in the detection of the microwave signal, it sometimes happens that the amplitude or phase unbalance of the bridge is not small, or that the changes in detected power are not small compared with the power itself. In these cases, the detected signals will not be pure χ' or χ'' signals, but rather will be distorted.

Consider a hybrid-T network (Fig. 57). The voltage in the receiving arm is:

$$\dot{V} = \dot{V}_1 + \dot{V}_2 ,$$

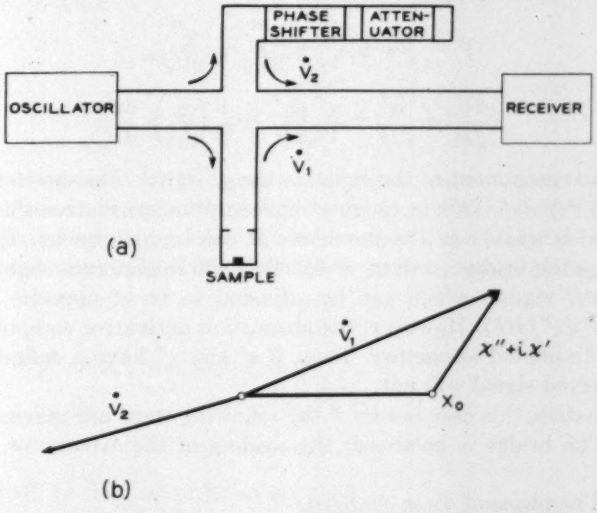


Fig. 57(a) — Outline of a hybrid bridge employed in electron spin resonance work; (b) vector diagram of the voltages entering the receiving arm of the bridge of (a).

where

$$\dot{V}_1 = x_0 + \chi'' + i\chi',\dagger$$
$$\dot{V}_2 = -x - iy,$$

with x_0, x, y, χ', χ'' being real numbers. (It is assumed that the cavity, or tuned circuit, is tuned to resonance.) Thus,

$$V = \sqrt{|\dot{V}_1 - \dot{V}_2|^2} = \sqrt{(x_0 + x + \chi'')^2 + (y + \chi')^2} \approx V_0 + \frac{x + x_0}{V_0}\chi'' + \frac{y}{V_0}\chi' + \frac{\chi''^2}{2V_0} + \frac{\chi'^2}{2V_0} - \frac{[(x + x_0)\chi'' + y\chi']^2}{2V_0^3} + \frac{[(x + x_0)\chi'' + y\chi']^3}{2V_0^5} + \cdots,$$

where

$$V_0 \equiv \sqrt{(x + x_0)^2 + y^2},$$

and it is assumed that

$$\chi'^2 + \chi''^2 < V_0^2.$$

Case 1

Let $x + x_0 = 0$. Then

$$V = V_0 + \chi' + \frac{\chi''^2}{2V_0} + \frac{\chi'^3}{2V_0^2},$$
$$\frac{dV}{dH} = \frac{d\chi'}{dH} + \frac{\chi''}{V_0}\frac{d\chi''}{dH} + \frac{3}{2}\left(\frac{\chi'}{V_0}\right)^2\frac{d\chi'}{dH}.$$

The main component of the signal is the χ' signal. The interference signal $(\chi''/V_0)(d\chi''/dH)$ is roughly of absorption-derivative shape, however, and is weaker at the shoulders. It can be reduced by slightly unbalancing the bridge ($x + x_0 \neq 0$). This will introduce a slight amount of $d\chi''/dH$ signal, which can be adjusted to be of opposite phase to $(\chi''/V_0)(d\chi''/dH)$. However, the absorption derivative component cannot be eliminated altogether. Thus, if χ' and χ'' have a definite parity, the observed signal will not.

In practice, this case occurs if the following steps are taken:

(a) The bridge is balanced; the reading of the attenuator, α db, is noted.

(b) The phase of V_2 is changed.

† Actually, the voltage is of course not equal to $\chi'' + i\chi'$, but only approximately proportional to it. For the other factors affecting it, see Ref. 6.

(c) The amplitude of V_2 is changed, so that a minimum detector output is observed, and the reading of the attenuator, β db, is noted.

(d) The amplitude of V_2 is changed by setting the attenuator at $2\alpha - \beta$ db. From Fig. 58, it is evident that, after step (a), $\dot{V}_2 = \mathbf{OA}$; after step (b), $\dot{V}_2 = \mathbf{OB}$; after step (c), $\dot{V}_2 = \mathbf{OC}$; and after step (d), $\dot{V}_2 = \mathbf{OD}$. From the geometry, $\overline{OB} = \overline{OA} = \sqrt{\overline{OC}\,\overline{OD}}$. Thus, $\log \overline{OD} = 2 \log \overline{OB} - \log \overline{OC}$, yielding our result.

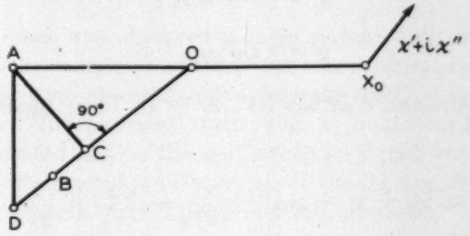


Fig. 58 — Vector diagram illustration of the voltages corresponding to cases (i), (iii) and (iv).

Case 2

Let $y = 0$. Then

$$V = V_0 + \chi'' + \frac{\chi'^2}{2V_0} + \frac{\chi''^3}{2V_0^2},$$

$$\frac{dV}{dH} = \frac{d\chi''}{dH} + \frac{\chi'}{V_0}\frac{d\chi'}{dH} + \frac{3}{2}\left(\frac{\chi''}{V_0}\right)^2 \frac{d\chi''}{dV_0}.$$

The main component of the signal is a χ'' signal. Again, there is an interference which can be reduced by slightly unbalancing the bridge ($y \neq 0$), but not eliminated. In practice, this case occurs if the bridge is balanced, and then the amplitude of V_2 is changed, without changing its phase.

Case 3

Let $x^2 + y^2 = x_0^2$, and say that

$$x = -x_0 \cos \varphi_0 ,$$

$$y = x_0 \sin \varphi_0 .$$

If φ_0 is small, to the first order in φ_0, then

$$V \approx V_0 + \chi' + \tfrac{1}{2}\varphi_0 \chi'' + \frac{\chi''^2}{2V_0} - \frac{\chi'\chi''}{4V_0} + \frac{\chi'^3}{2V_0^2}.$$

This case is obtained when the bridge is balanced, and then the phase of V_2 is changed, but not its amplitude. We have a "χ'" signal, with some "χ''" admixed.

Case 4

Let

$$x = -r \cos \varphi_0 ,$$

$$y = r \sin \varphi_0 .$$

Let φ_0 be constant and r satisfy $\partial V_0/\partial r = 0$. Then, again to the first order in φ_0,

$$V \approx V_0 + \chi' + \varphi_0 \chi'' + \frac{\chi''^2}{2V_0} - \frac{\chi'\chi''}{V_0} + \frac{\chi'^3}{2V_0^2} .$$

This case is obtained when the bridge is balanced, then the phase of V_2 is changed, and then the amplitude of V_2 is changed so that minimum microwave power is detected. Again, we have a χ' signal with a small admixture of χ''.

We see that the distortions of the χ' and χ'' signals are of two kinds:

i. Those due to a nonvanishing φ_0, i.e., to a large unbalance of the bridge.

ii. Those due to nonvanishing χ'/V_0, χ''/V_0, i.e., to "strong" signals. The distortions of the first kind can be removed altogether by proper adjustment of the bridge. (If $V_0 \gg \chi', \chi''$, Case 1 above will yield a pure χ' signal; while Case 2 will yield a pure χ'' signal.)

The distortions of the second kind cannot be removed by any adjustment of the bridge. They will give rise to errors in measurements of the center of the line, its width and its intensity.

Therefore, if V_0 is not large compared with χ', χ'' (a condition that is indicated by a change in detector crystal current on passing through line), it is recommended to *increase* the detuning of the bridge, and if a χ' signal is desired, manipulate it as described in Case 1 of this Appendix. (It may then be necessary to introduce attenuation in the bridge output arm to avoid overloading the detector.)

An experiment was carried out to check these calculations, employing as bridge elements a Hewlett-Packard X885A phase shifter and a Hewlett-Packard X382A attenuator. The sample used was D.P.P.H., and the experiments were performed at room temperature.

The experimental results agree with these calculations, and indicate that the above bridge elements can be assumed to be ideal. A quite pure χ' signal could be obtained by the method indicated in Case 1, even for a phase shift $\varphi_0 = 60°$.

APPENDIX F

A Particular χ' Trace Observed at High Powers

A peculiar trace was observed at high powers and small dH/dt with the 0.37 ohm-cm sample. The trace (as observed at the output of the IF amplifier, with no phase sensitive detector used) is shown in Fig. 59. It looks like a "slow passage" trace (Fig. 2) but, after passing the line once, it is saturated and on the next sweep no signal whatsoever is seen. If the trace were due to slow passage, it should not depend upon the history of the system, and thus should not saturate.

A possible explanation for this phenomenon is as follows: At high powers, forbidden lines are excited. These forbidden lines correspond to a simultaneous flip of an electron and a Si^{29} nucleus interacting with it. These forbidden lines have been observed directly and are discussed in a separate paper. Now, each donor interacts with many Si^{29} nuclei and thus gives rise to many forbidden lines. Under the given experimental conditions each forbidden line is traversed "almost suddenly." Thus, the superposition of these lines should give rise to a dispersion signal (Appendix B). The probability of a donor to flip during the passage of each forbidden line is small, since the passage is almost sudden. But, since there are many forbidden line in which each donor participates, the over-all probability to flip may be considerable, and we may obtain saturation.

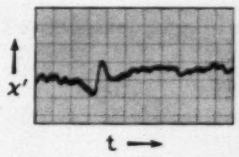


Fig. 59 — χ' of the bound donor electron line in phosphorus doped silicon, under strong microwave power, with no phase sensitive detector (or magnetic field modulation) present. The shape of the line resembles that of Bloch's "slow passage" case; however, the line saturates after a single passage through it. Data: Sample, phosphorus-doped silicon; impurity concentration, 1.7×10^{16} atoms cc; $T = 1.25°K$; $T_1 \approx T_2 \approx 20$ seconds; $H \approx 3200$ gauss; $H_1 \approx 1/30$ gauss; $dH/dt = 3$ gauss/second.

REFERENCES

1. Bloch, F., Phys. Rev., **70,** 1946, p. 460.
2. Bloembergen, N., Purcell, E. M. and Pound, R. V., Phys. Rev., **73,** 1948, p. 679.
3. Jacobsohn, B. A. and Wangness, R. K., Phys. Rev., **73,** 1948, p. 942.
4. Salpeter, E. E., Proc. Phys. Soc., **63A,** 1950, p. 337.
5. Portis, A. M., *Magnetic Resonance in Systems with Spectral Distributions*, A.R.D.C., 1955.
6. Feher, G., B.S.T.J., **36,** 1957, p. 449.
7. Redfield, A. G., Phys. Rev., **98,** 1955, p. 1787.
8. Wangness, R. K. and Bloch, F., Phys. Rev., **89,** 1953, p. 728.
9. Bloembergen, N., Shapiro, S., Pershan, P. S. and Artman, J. O., Phys. Rev., **114,** 1959, p. 445.
10. Pines, D., Bardeen, J. and Slichter, C. P., Phys. Rev., **106,** 1957, p. 489.
11. Portis, A. M., Phys. Rev., **91,** 1953, p. 1071.
12. Feher, G., Phys. Rev., **114,** 1959, p. 1219; Feher, G. and Gere, E. A., Phys. Rev., **114,** 1959, p. 1245.
13. Zener, C., Proc. Roy. Soc., **A137,** 1932, p. 696.
14. Gabillard, R., Comp. rend., **233,** 1951, p. 33.
15. Gordon, J. P. and Bowers, K. D., Phys. Rev. Letters, **1,** 1958, p. 368.
16. Abragam, A., to be published.
17. Jahnke, E. and Emde, F., *Tables of Functions*, Dover Publications, New York, 1945, p. 20.

Recent Monographs of Bell System Technical Papers Not Published in This Journal*

ANDERSON, E. W. and MCCALL, D. W.

Dielectric Properties of Linear Polyamides, Monograph 3530.

ANDERSON, O. L.

The Debye Temperature of Vitreous Silica, Monograph 3579.

ANDERSON, P. W.

Theory of Dirty Superconductors, Monograph 3541.

ANDERSON, P. W. and SUHL, H.

Spin Alignment in the Superconducting State, Monograph 3542.

BALLHAUSEN, C. J., see Liehr, A. D.

BOND, W. L., see Kaiser, W.

BOYLE, W. S.

Far Infrared Magneto-Optic Effects from Impurities in Germanium, Monograph 3502.

BRADY, G. W.

Cluster Formation in Perfluoroheptane, Monograph 3524.

BRATTAIN, W. H.

Surfaces, Monograph 3501.

BURRUS, C. A.

Stark Effect at 0.93, 1.18 and 1.5 Millimeters Wavelength: DCl, DBr and DI, Monograph 3485.

* Copies of these monographs may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. The numbers of the monographs should be given in all requests.

CHYNOWETH, A. G.

Radiation Damage Effects in Ferroelectric Triglycine Sulfate, Monograph 3271.

CHYNOWETH, A. G. and MCKAY, K. G.

Light Emission and Noise Studies of Microplasmas in Silicon p-n Junctions, Monograph 3486.

CLOSSON, H. T., see Melroy, D. O.

COMPTON, K. G.

Instruments for Measurements in Underground Corrosion Work, Monograph 3448.

COMPTON, B. V., see Matthias, B. T.

CORENZWIT, E., see Matthias, B. T.

CURTIS, H. E.

Probability Distribution of Noise Due to Fading on Multisection FM Microwave, Monograph 3525.

DE BUSKE, J. J., see Looney, D. H.

DEEG, G., see Leutritz, J., Jr.

DEVLIN, G. E., see Schawlow, A. L.

DILLON, J. F., JR. and EARL, H. E.

Demonstration of Magnetic Domains, Monograph 3295.

DILLON, D. M., see Egerton, L.

EARL, H. E., see Dillon, J. F., Jr.

EARLY, J. M.

Maximum Rapidly-Switchable Power Density in Junction Triodes, Monograph 3533.

EGERTON, L. AND DILLON, D. M.

Piezoelectric and Dielectric Ceramics in the System Potassium-Sodium Niobate, Monograph 3449.

FEHER, G.

Donor Wave Functions in Silicon by the ENDOR Technique, Monograph 3503.

FLOOD, W. F., see Haynes, J. R.

FOWLER, A. D. and IGLEHEART, J. D.

Effects of Frequency Cutoff Characteristics on Spiking and Ringing of TV Signals, Monograph 3534.

FRISCH, H. L., HELLMAN, M. Y. and LUNDBERG, J. L.

Adsorption of Polymers: Polystyrene on Carbon, Monograph 3450.

FRISCH, H. L., see Pollak, H. O.

FRISCH, H. L., see Reiss, H.

FROST, H. B.

New Methods for the Measurement of Cathode Interface Impedance, Monograph 3535.

GEBALLE, T. H., HERRING, C. and KUNZLER, J. E.

Phonon-Drag Thermomagnetic Effects in N-Germanium, Monograph 3504.

GILBERT, E. N.

Random Graphs, Monograph 3488.

GOLDSTEIN, H. L.

Transients in Series-Connected Saturable Reactor with High-Impedance Control, Monograph 3527.

HAGSTRUM, H. D.

State Density in the Valence Band of Silicon, Monograph 3505.

HAYNES, J. R., LAX, M. and FLOOD, W. F.

Analysis of Intrinsic Recombination Radiation from Silicon and Germanium, Monograph 3506.

HEBEL, L. C.

Theory of Nuclear Spin Relaxation in Superconductors, Monograph 3489.

HELFAND, E., see Reiss, H.

HELLMAN, M. Y., see Frisch, H. L.

HERRING, C.

Transport, Monograph 3507.

HERRING, C., see Geballe, T. H.

HOLBROOK, B. D.

Some Logical Requirements for the Control of Switching Networks, Monograph 3520.

HOWARD, J. B. and MARTIN, W. M.

Effects of Thermal History on Some Properties of Polyethylene, Monograph 3577.

HUMPHREY, F. B., REYNOLDS, F. W. and STILWELL, G. R.

Introduction to Magnetic Thin Films, Monograph 3431.

HUTSON, A. R.

Electronic Properties of ZnO, Monograph 3508.

IGLEHEART, J. D., see Fowler, A. D.

JANIK, J., JR., see Looney, D. H.

KAC, M. and SLEPIAN, D.

Large Excursions of Gaussian Processes, Monograph 3490.

KAMINSKY, G., see Lee, C. A.

KAISER, W. and BOND, W. L.

Nitrogen, A Major Impurity in Common Type 1 Diamond, Monograph 3455.

KARNAUGH, M.

Magnetic Selectors, Monograph 3521.

KOEHLER, D. C.

Automatic Processing of Photographic Code Plates for Electronic Telephone Switching, Monograph 3491.

KUMMER, W. H.

Sweep-Frequency Studies in Beyond-the-Horizon Propagation, Monograph 3526.

KUNZLER, H. E., see Geballe, T. H.

LAX, M.

Fluctuations from the Nonequilibrium Steady State, Monograph 3543.

LAX, M.

Giant Traps, Monograph 3509.

LAX, M., see Haynes, J. R.

LEBOWITZ, J. L., see Reiss, H.

LEE, C. A. and KAMINSKY, G.

Temperature Variation of Noise in Diode and Transistor Structures, Monograph 3528.

LEENOV, D. and UHLIR, A., JR.

Generation of Harmonics and Subharmonics with p-n Junction Diodes, Monograph 3578.

LEWIS, W. D.

Microwave Logic, Monograph 3522.

LEUTRITZ, J., JR., MCMAHON, W. AND DEEG, G.

Relationship of Evaporation Pattern and Distillation of Coal Tar Creosote, Monograph 3456.

LEUTRITZ, J., JR., see Olafson, P.

LIEHR, A. D.

Reciprocation of Electrostatic and Electromagnetic Forces in Ligand Field Theory, Monograph 3529.

LIEHR, A. D. and BALLHAUSEN, C. J.

Complete Theory of Ni(II) and V(III) in Cubic Crystalline Fields. Octahedral Complexes, Monograph 3559.

Logan, R. A. and Peters, A. J.

Diffusion of Oxygen in Silicon, Monograph 3492.

Looney, D. H., De Buske, J. J., Janik, J., Jr. and Simons, B. H.

Twistor Matrix Memory; Card Changeable Nondestructive Readout Store, Monograph 3406.

Lundberg, J. L., see Frisch, H. L.

Mandell, E. R., see Slichter, W. P.

Martin, W. M., see Howard, J. B.

Matthias, B. T., Compton, V. B., Suhl, H. and Corenzwit, E.

Ferromagnetic Solutes in Superconductors, Monograph 3536.

Melroy, D. O. and Closson, H. T.

Measurement of Internal Reflections in Traveling-Wave Tubes, Monograph 3537.

McCall, D. W., see Anderson, E. W.

McKay, K. G., see Chynoweth, A. G.

McMahon, W., see Leutritz, J., Jr.

McSkimin, H. J.

High-Frequency Barium Titanate Transducers for Generating Ultrasonic Waves in Liquids, Monograph 3458.

Miller, R. C., see Savage, A.

Moore, E. F.

The Shortest Path Through a Maze, Monograph 3523.

Moore, G. E.

Dissociation of Solid SrO by Impact of Slow Electrons, Monograph 3434.

Morgan, S. P.

Generalizations of Spherically Symmetric Lenses, Monograph 3531.

Moriya, T.

Piezomagnetism in CoF_2, Monograph 3545.

OLAFSON, P. and LEUTRITZ, J., JR.

The Toxicity of Creosote and Creosote-Pentachlorophenol Mixtures to Cattle, Monograph 3461.

OLMSTEAD, P. S.

Production and Statistics, Monograph 3540.

OLMSTEAD, P. S.

Quality Control and Operations Research, Monograph 3544.

ORTEL, W. C. G.

Nanosecond Logic by Amplitude Modulation at X-Band, Monograph 3462.

PETERS, A. J., see Logan, R. A.

PHILLIPS, J. C.

Energy-Band Interpolation Scheme Based on a Pseudopotential, Monograph 3276.

POLLAK, H. O. and FRISCH, H. L.

The Time Lag in Diffusion, Monograph 3426.

REISS, H., FRISCH, H. L., HELFAND, E. AND LEBOWITZ, J. L.

Aspects of the Statistical Thermodynamics of Real Fluids, Monograph 3532.

REISS, H.

Diffusion-Controlled Reactions in Solids, Monograph 3465.

REYNOLDS, F. W., see Humphrey, F. B.

RINEY, T. D.

Coefficients in Certain Asymptotic Factorial Expansions, Monograph 3468.

SAVAGE, A. and MILLER, R. C.

Asymmetric Hysteresis Loops and the Pyroelectric Effect in Triglycine Sulfate, Monograph 3494.

SCHAWLOW, A. L. and DEVLIN, G. E.

Effect of Energy Gap on Penetration Depth of Superconductors, Monograph 3277.

SCHEINMAN, A. H.

Numerical-Graphical Method in the Design of Multiterminal Switching Circuits, Monograph 3495.

SIMONS, B. H., see Looney, D. H.

SLEPIAN, D., see Kac, M.

SLICHTER, W. P. and MANDELL, E. R.

Molecular Motion in Some Glassy Polymers, Monograph 3470.

SMITH, D. H.

One-Watt Solar Power Plant, Monograph 3467.

SPECTOR, C. J.

A Design Theory for the High-Frequency p-n Junction Variable Capacitor, Monograph 3538.

STILWELL, G. R., see Humphrey, F. B.

SUHL, H.

Note on the Saturation of the Main Resonance in Ferromagnetics, Monograph 3498.

SUHL, H., see Anderson, P. W.

SUHL, H., see Matthias, B. T.

THOMAS, D. G.

Infrared Absorption in Zinc Oxide Crystals, Monograph 3499.

UENOHARA, M.

Noise Consideration of the Variable Capacitance Parametric Amplifier, Monograph 3539.

UHLIR, A., JR., see Leenov, D.

WEINREICH, G.

Chemical Shift of Donor States in Germanium, Monograph 3511.

Contributors to this Issue

ANDREW J. AIKENS, B.S.(E.E.), 1920, University of Nevada; Pacific Telephone and Telegraph Company, 1922–28; American Telephone and Telegraph Company, 1928–34; Bell Telephone Laboratories, 1934—. Mr. Aikens has specialized in studies of noise and noise prevention in telephone plant. He now supervises a group engaged in work on interference problems.

W. J. ALBERSHEIM, E.E., 1922, and Eng.D., 1924, Technical Institute, Aachen, Germany; E.R.P.I., 1929–41; Bell Telephone Laboratories, 1941–59. With E.R.P.I., Mr. Albersheim was engaged in studies of sound recording and reproducing methods. After joining Bell Laboratories he worked in radio research, including radar and countermeasures during World War II, and later FM theory and studies of microwave systems for multiplex and long-distance voice and television channels. From 1953 to 1958 he supervised a group of specialists in military problems including radar, countermeasures and missile guidance. At the time of his retirement in 1959 he was charged with organizing a computing center for the Whippany Laboratories. At present he is vice president for engineering of the Spencer-Kennedy Laboratories, Boston. Fellow I.R.E.; member New York Academy of Sciences, S.M.P.T.E.

M. M. ATALLA, B.S., 1945, Cairo University; M.S., 1947, and Ph.D., 1949, Purdue University; Bell Telephone Laboratories, 1950—. For five years Mr. Atalla headed a group engaged in basic studies in contact physics. He is now in charge of a group doing studies on the physics of semiconductor surfaces and fundamental development of semiconductor devices. Member American Physical Society, Sigma Xi, Sigma Pi Sigma, Pi Tau Sigma.

P. G. BHUTA, B.Sc., 1951, University of Bombay; B.S.E., 1954, University of Michigan; M.S., 1957, Pennsylvania State University; Bell Telephone Laboratories, 1957—. He has been engaged in structural design and research in applied mechanics. Member Chi Epsilon, Tau Beta Pi.

WILLIAM T. COCHRAN, B.S.E.E., 1956, Purdue University; Bell Telephone Laboratories, 1956—. His early assignments included development of a highly accurate oscillator for a measuring set, an echo suppressor for TASI, and a system for measuring phase of complex waveforms. He has most recently been engaged in design and development of a noise measuring set. Member Eta Kappa Nu, Sigma Pi Sigma, Tau Beta Pi.

SIDNEY DARLINGTON, B.S., 1928, Harvard College; B.S. in E.E., 1929, Massachusetts Institute of Technology; Ph.D., 1940, Columbia University; Bell Telephone Laboratories, 1929—. He has been engaged in research in applied mathematics with emphasis on network theory and military electronics. He holds more than 20 patents in these fields. Fellow I.R.E.; member American Rocket Society.

RICHARD L. DEININGER, B.S., 1950, and M.S., 1952, Carnegie Institute of Technology; Ph.D., 1955, Ohio State University; U.S. Air Force Aeromedical Laboratory, Wright Air Development Center, 1955–58; Bell Telephone Laboratories, 1958—. He has been engaged in studies of human factors relating to pushbutton telephone design and other telephone services. Member American Statistical Association, Society of Engineering Psychologists.

A. ROSS ECKLER, B.A., 1950, Swarthmore College; Ph.D., 1954, Princeton University; Bell Telephone Laboratories, 1954—. Since joining Bell Laboratories, Mr. Eckler has served as a mathematical consultant, in the fields of statistics and probability. He supervises a group of mathematical consultants who are concerned with military projects, systems studies and mathematical research. Member American Statistical Association, Institute of Mathematical Statistics, Society for Industrial and Applied Mathematics, Phi Beta Kappa, Sigma Xi.

JAMES GAMMIE, B.Sc.(E.E.), 1944, University of Aberdeen; B.Sc., (mathematics), 1951, Birkbeck College, University of London; Standard Telephones and Cables, Ltd. (England), 1944–51; Bell Telephone Laboratories, 1952—. He has been engaged in development work on transmission systems, including the L-3 coaxial system and, more recently, short-haul microwave radio relay systems.

S. D. HATHAWAY, B.E.E., 1947, University of Virginia; M.S.E.E., 1950, Virginia Polytechnic Institute; faculty, Virginia Polytechnic Insti-

tute, 1947–50; M.S.E.E., 1952, University of Illinois; Bell Telephone Laboratories, 1952—. He has been engaged in systems engineering on microwave radio relay systems, including studies of the effects of rainfall on radio transmission. He supervises a group working on light-route systems. Member I.R.E., Eta Kappa Nu, Tau Beta Pi.

JOHN R. KLAUDER, B.S., 1953, University of California; M.S., 1956, Stevens Institute of Technology; M.A., 1957, and Ph.D., 1959, Princeton University; Bell Telephone Laboratories, 1953—. Mr. Klauder first worked on design of impedance measuring equipment for traveling wave tubes and studies of advanced radar designs. He was a member of the Communications Development Training Program from 1953 to 1956, and from 1956 to 1959 was a C.D.T. Fellow at Princeton, where he specialized in quantum field theory. Since returning to Bell Laboratories, he has been engaged in theoretical studies of impurity band properties and the fundamental properties of metals. Member American Physical Society, Phi Beta Kappa, Sigma Xi, Tau Beta Pi.

DAVID A. LEWINSKI, B.Sc., 1952, and M.Sc., 1956, University of Manitoba; Bell Telephone Company of Canada, 1954–56; Bell Telephone Laboratories, 1956—. At Bell Laboratories he has been engaged in interference studies, with emphasis on the subjective aspects of noise prevention. Member Mathematical Association of America.

AMOS C. PRICE, JR., B.S.E.E., 1953, Swarthmore College; Bell Telephone Laboratories, 1953—. He has been engaged in development and analysis of experimental radar systems. In 1958–59 he was an instructor in a Communications Development Training Program course in radar systems engineering. Member I.R.E., Sigma Xi.

IRWIN W. SANDBERG, B.E.E., 1955, M.E.E., 1956, and D.E.E., 1958, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1958—. He has been concerned with analysis of military systems, particularly radar systems, and with synthesis and analysis of active and time-varying networks. Member I.R.E., Eta Kappa Nu, Sigma Xi, Tau Beta Pi.

EILEEN TANNENBAUM, B.A., 1950, and M.A., 1952, Mount Holyoke College; Ph.D., 1955, University of California; Bell Telephone Laboratories, 1956—. Miss Tannenbaum has been engaged in work on solid state devices and basic research in surface studies of semiconductors, especially with regard to transistors. Member American Physical Society.

MEIR WEGER, M.S., 1957, Hebrew University (Jerusalem); Bell Telephone Laboratories, Summer, 1959. Mr. Weger is studying toward the Ph.D. degree at the Univesity of California. He has been concerned with electron spin resonance and nuclear magnetic resonance in solids. Member British I.R.E.